Table of Contents

Table of Contents

VEHICLE ROUTING APPROACH FOR LEAN SHOP-FLOOR LOGISTICS

DAVID GYULAI[1,2][✉] AND LÁSZLÓ MONOSTORI[1,2]

[1] Fraunhofer Project Centre for Production Management and Informatics, Computer and Automation Research Institute (SZTAKI), Kende str. 13–17, Budapest, H-1111, HUNGARY
[2] Department of Manufacturing Science and Technology, Budapest University of Technology and Economics, Egry J. str. 1, Budapest, H-1111 HUNGARY
✉ E-mail: david.gyulai@sztaki.mta.hu

In order to satisfy the material supply needs of large scale shop-floors and production systems, various logistics solutions are applied. In lean manufacturing enterprises, the material supply is pulled by the demands of manufacturing/assembly processes; therefore, a milkrun service is often applied to support the production without glitches. The milkrun logistics planning is a special case for vehicle routing problem (VRP), and requires effective approach to solution in order to satisfy various constraints, and minimize the cost of service. This study gives an overview about lean logistics as well as the most efficient VRP solver algorithms. Furthermore, a novel initial solution with generation heuristics is proposed, which is specially focused on flexible milkrun planning. In order to demonstrate the capabilities of the solution, a software environment is developed as a demonstration that focuses on the main industrial requirements of logistics planning like effective layout definition, quick response of the delivery service and effective order handling.

Keywords: vehicle routing problem, milkrun, local search, factory logistics

Introduction

The material provision of large- and medium-scale production systems requires prudent planning and control, since it influences the performance of the production and affects the order management directly and storage assignment [1]. In order to manage effectively the material provision of the production, practical transport logistics and distribution tasks are usually formulated as vehicle routing problems, whose objective is to obtain a minimum-cost route plan serving a set of customers with known demands [2].

In state-of-the art manufacturing systems, lean processes are successfully applied to eliminate the wastes on all the levels of production, and they raise the overall efficiency of the system by reducing the non-value adding activities in the process chain. The shop-floor logistics processes have to be adapted to the production system; therefore the waste reduction in the material provision also has a key role in balancing the workload and ensuring a smooth running of production. A well-operating lean logistics system keeps the inventories and operational costs on the most cost efficient levels, via by providing the materials what that are needed, exactly when they are needed. For satisfying these requirements, concepts of the lean logistics can be applied that are primarily aimed at controlling the transport services duly to the pull strategy.

The paper is focused on the 'milkrun' service, which is defined as a manually operated, cyclic transport system delivering raw materials (or sub-assemblies) and disposing empties based on consumption using a fixed route and time schedule [3]. According to this characterization, milkrun scheduling is considered as a special vehicle routing problem with time windows, a limited capacity, and number of vehicles. In this paper a novel, multi-level planning approach is proposed, which combines the advantages of existing algorithms as for example local search for vehicle routing.

First, an overview of the state-of-the-art factory logistics is given, considering the recent shop-floor characteristics and planning demands. Then the main solution approaches and algorithms are introduced, as well as the critical points of milkrun planning. In section 4, the proposed layout representation hierarchy and the novel initial solution generation heuristics method are described. The latter focuses on the real industrial criteria and give a feasible milkrun schedule. In the last section, an industrial case-study and some numerical results present the efficiency of the milkrun planning method developed.

Shop-floor logistics

The shop floor logistics are part of the production logistics in the supply chains of companies. While logistics in general is about supplying a customer with

necessary goods in time, with the goal of efficiency, and the lowest possible costs, production logistics is the task of bringing necessary material to workstations and machines and transferring it to the next station. In the realm of logistics planning methodologies the lean approach, originating in the Toyota Production System, of overall waste reduction has become a major planning paradigm, constituting the lean logistics approach [4].

Principles of lean logistics

Of the seven waste categories, eliminating unnecessary transports, waiting times, superfluous movement and excess inventory are the most relevant in lean logistics. KLUG identifies the following characteristics of lean logistics [5]: synchronicity and a clocked material flow, flow orientation, standardization, pull principle, stability, integration, as well as perfection. Regarding the pull principle, as the most relevant feature, material is only transferred to a workstation if it is about to process the material thus reducing inventory levels by reducing shop floor stock to safety buffers only. Beside the above principles of lean logistics, a relatively novel issue in material provision of production is 'green logistics' that is generally about establishing an ecologically sustainable transportation system considering factors like energy consumption and air-pollution [7].

While in the general external logistics context theses planning principles are reflected in transport concepts such as the just-in-time delivery, there are also corresponding design concepts in production and shop floor logics. For the supplier/source side, establishing zones for decentralized incoming goods and supermarkets in the vicinity of assembly stations is a means of reducing transport lengths. On the receiving side, standardized shelves and reusable containers, replenishment schemes like 'kanban' are applied. For the transport system between supply and workstations, 'tugger trains' *(Fig.1)* with fixed time intervals, standardized capacity, and transport lot sizes are preferred over forklift transport in the context of lean logistics; the goal of, which is the reduction of inventory and avoiding erratic peaks in production known as the bullwhip effect [8].

This transport concept of scheduled tugger trains in the context of a lean-oriented logistics system is usually combined with the milkrun transport concept that is in the focus of this paper. Especially in production environments with a high product complexity, variety amid a complex and changing material flow, planning a system of milkrun transports is a complex planning task. Thus, an efficient system of milkruns could significantly in-crease the efficiency of the overall production logistics.

Milkrun *as a vehicle routing problem*

Vehicle routing (VRP) is a challenging combinatorial optimization problem, introduced first by DANTZIG and RAMSER in 1959 [9]. The general VRP is known as the Capacitated Vehicle Routing Problem (CVRP), were the aim is to satisfy the needs of all the customers at different locations by having a given number of vehicles with capacity constraints [11]. When milkrun is applied as a material providing system, the time constraint of the runs has also to be added to the general problem (VRP with Time Windows – VRPTW). The general objective is to minimize the costs either by minimizing the total distance travelled, or minimizing the number of vehicles applied. A specific type of VRP is the Pickup and Delivery Problem (PDP) where the vehicles not only provide the locations with materials, but also pick up materials at the stations and deliver them to another one. The milkrun problem is also a PDP, with time windows and cyclic service [12].

Several approaches from the operations science are used to solve VRPs, as for example constraint programming that manages flexibly the various specific constraints of the VRP [14]. Local search is one of the fundamental approaches to find solutions for hard computational – including constraint satisfaction – problems. The basic idea behind local search is to start with a randomly or heuristically generated candidate solution, and to iteratively improve that by means of defined functions [15].

Generating the initial solution is always a crucial point when applying local search, since it significantly affects the running time of the algorithm. Sophisticated heuristics for the VRP are the two-phased ones, which decompose the problem into a clustering- and a routing problem, with possible feedback loops between the two stages. Cluster-First Route-Second (CFRS) algorithms perform a single clustering of the vertex set and then they determine a vehicle route on each cluster. The best-known CFRS algorithm is the FISHER-JAIKUMAR algorithm, which solves the General Assignment Problem (GAP) to form the clusters [17].

The local search starts from the initial solution and subsequently moves from the present solution to a neighbouring solution in the search space where each solution has only a relatively small number of feasible neighbour solutions and each movement is determined by neighbouring operators [20]. When applying local search, the combination of the neighbourhood operators produces the next local optimum solution *(Fig.1)*.

Although the most general forms of the VRP can be solved effectively with the above approaches, the milkrun planning problem is hard to interpret as a simple graph-search problem, and the general initial solution heuristics are difficult to transform for this problem. The mapping process of the factory layout also requires special processing; therefore general solution methods cannot be applied easily for such a special problem.
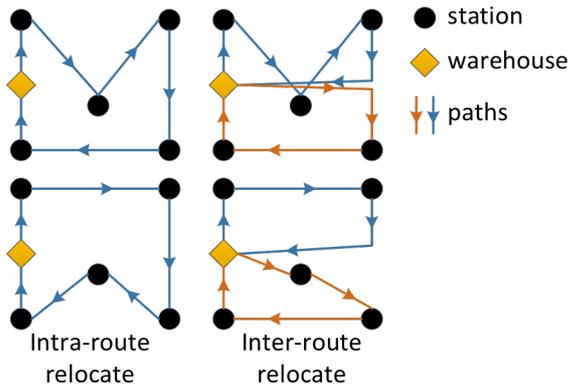
Figure 1: Inter- and intra-route relocate operators

Figure 2: Layout elements and vehicle movement

Problem formulation

Hereby, we provide a definition to the milkrun planning problem. Consider a layout of the shop-floor, which is defined by a set of routes and stations. The routes can be either one- or two-way ones, and each route endpoint is defined as a "routenode". Along a specific route, there can be more stations *(Fig.2)*. The problem also considers demands that belong to the stations and defined by the amount of the transported goods in standard units and the required cycle-times of the transportation. The goods are transported by vehicles (practically by tugger trains), which are defined by their capacity and average speed.

In order to plan feasible milkruns, real-world constraints are considered, such as the capacity constraints of the vehicles, which limit the maximum number of the transported goods. Another constraint is the time limit of the plan, which means the total time consumption of a milkrun plan cannot exceed the cycle-time of the demands. The loading and unloading points can be reached from both sides of the tugger train, and the train can approach the station from both endpoints of the route that the station belongs to.

A milkrun plan is built-up by paths, where each path is given as the list of the visited stations and the list of the routes that the vehicle passes along. The plan does not partition the demands, and a visit of a particular station only occurs in a specific path. A milkrun plan is characterized by its total time consumption, which is required by the vehicle(s) to perform the plan. A milkrun plan is considered to be better than the other one only if its total time consumption is smaller, while it satisfies all the demands. The goal of milkrun planning is to minimize the number of the required vehicles via minimizing the time consumption of the plan.

The proposed planning process

Layout representation for vehicle routing

As it was mentioned in the previous section, solving the general VRP is not enough to plan feasible milkrun schedule, since most of the routing algorithms does not calculate with the physical constraints determined by the shop-floor. Most of the general solution approaches consider the problem as a graph-search, where the stations are represented by vertices and the routes are represented by edges [21].
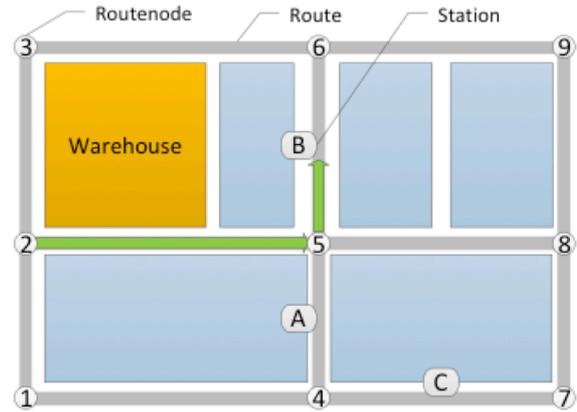
In many cases, this representation does not support effectively the planning processes with feasible results, unless the application of the following constraints. Frequently, when representing a shop-floor by directed graphs, it turns out that the structure of the layout results some direct and indirect routing constraints. A direct case means that the additional equations can be formalized immediately when constructing the graph, such as the asymmetric edge formulas, while the others require further consideration. Indirect constraints are usually implied by the narrow and one-way corridors, which limit the abilities of the vehicles from going to a station directly from another, even if a path connects them. These constraints are implied by the moving abilities of the common applied tugger train, since it cannot turn in the middle of an aisle, and in many of the cases it have to take a detour before visiting the next station *(Fig.2)*. In order to handle the shop-floor constraints effectively and be able to avoid impossible movements of the vehicle, a novel hierarchical layout representation is proposed. The layout is defined by three main different classes, each having their own specific attributes. This structure can be handled dynamically during the path calculations, and the asymmetric nature of the distance matrix is represented together with the limitations of vehicles movements. *Fig.3* shows the hierarchy of the layout, defined by "routenodes", "routes" and "stations".

To calculate the distance matrix of the layout, storing the shortest paths between the stations, DIJKSTRA'S algorithm was applied that solves the single-source shortest path problem in logarithmic running time [22]. In this case, the input of the algorithm is a directed graph where the vertices are the set of the "routenodes" and the edges are the routes. This graph representation is able to handle all the nodes of the layout with their connections, and by applying DIJKSTRA'S algorithm the asymmetric distance matrix can be calculated.
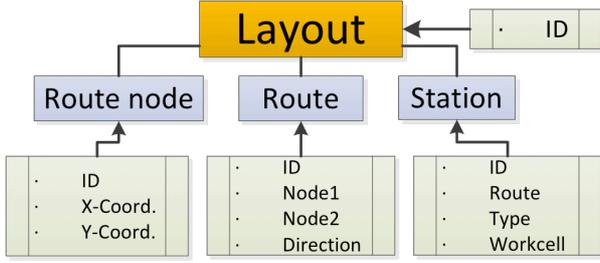
Figure 3: Hierarchy of the layout representation

Initial solution generation

In order to plan feasible milkrun cycles with the lowest operational costs, the shop-floor constraints have to be taken into consideration. To handle the constraints effectively, the routing algorithm needs to select the stations to be visited dynamically. Thus, a so-called triplet solution is applied, which helps to avoid impossible movements of the vehicle. A triplet is consisted of three identical nodes: three current, the previous and the next position of the vehicle. The next position is calculated dynamically when determining a particular path of the vehicle. The milkrun planning method strongly focuses on the industrial requirements of vehicle routing; therefore, a novel initial solution generation heuristics was implemented. The goal is to generate a solution that is as close to the criterion as possible.

The proposed heuristics can be classified as a CFRS type rule, and differs from the FISHER-JAIKUMAR algorithm by the cluster generation method, since it does not require solving time-consuming GAP, but defines the clusters based on practical reasons [18].

To generate a feasible initial solution, the algorithm calculates a path to each station applying DIJKSTRA'S shortest path algorithm. Then a greedy search algorithm is applied to calculate the next node, which must be visited by the vehicle before returning to the depot, so as to avoid violating the turnaround constraints of the vehicle. The greedy search iterates forward the nodes applying a best-first search strategy [23], and finds the first node from, which the shortest path to the depot does not contain the previous node. The "roundtrip" planning method is applied for each station, and detects all the feasible tours on the shop floor. In order to determine the set of paths for the initial solution, the paths with most visits are selected one by one while the set of unsatisfied demands is not void. Performing the services required by these cycles provides a feasible and acceptable initial solution for the planning algorithm.

The pseudo-code of the proposed algorithm is the following:

$v \in V$	set of stations
$p := \{v \mid v \in V\}$	path
$P\{p\}$	set of paths
$D = [d_{rs}] \mid (r, s) \in V$	distance matrix

$I\{p\}$	initial solution
q_i	demands of the ith station
$Q = \{q_i\}$	set of demands
t_c	cycle time
c	capacity of the vehicle

initialize *V, I, Q, C{v}, c, tc*
foreach *i* in *V*
    `calculate` D_{0i}
    foreach *v* in D_{0i}
        `add` *v* to C_i
    `calculate` D_{i0}
    foreach *v* in D_{i0}
        `add` *v* to C_i
`align` *C* [`length`(C_i+1) < `length` (C_i)]

while *|Q| > 0*
    while `time`(C_l) > *tc* or `load`(C_l) > *c*
        `remove` rand *v* from C_l (*v≠i*)
    foreach *v* in C_l
        `delete` q_v from *Q*
    `add` C_l to *I*
    `delete` C_l from *C*
return *I*

Local search strategy and neighbourhood functions

In order to improve the initially generated solution, neighbourhood functions are applied. To be able to reach all the points of the search space, both intra- and inter-route operators are necessary to calculate with, although in this case the effect of the inter-route functions is more significant since the initial solution heuristics could provide paths with stations in a rather good sequence. The algorithm performs the local search to minimize the paths, and the cycle-time of them. Then, a simple bin-packing problem is considered, to allocate the paths to a minimal number of vehicles. To solve the problem, first fit heuristics was applied, which is a proven 11/9 OPT solution algorithm [24].

Case study and test results

The milkrun planner software, which includes both the user interface and the solver algorithms were implemented in a .NET environment using C++ language. Within the performance evaluation of the implemented software, a real production environment was modelled in order to analyse its capabilities. The purpose of the implemented application is to offer user-friendly and efficient milkrun planning environment; therefore a graphical, point-and-click layout definition interface and XML communication have been applied *(Fig.4)*.
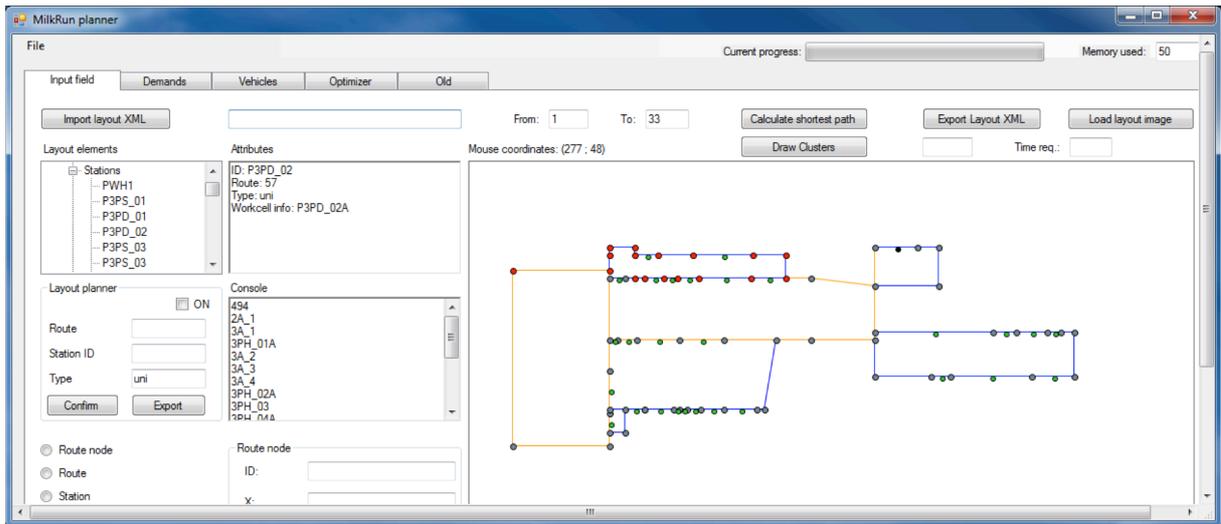
Figure 4: The user interface of the milkrun planner application

First the input parameters are processed and stored in the memory, and DIJKSTRA'S algorithm is applied to calculate the shortest path between each station. The distance matrix is built up from vectors, each having two distance elements. This data structure is required by the applied triplet-based path planning method, which selects dynamically the proper distance parameter from the vector, based on the location of the previously visited station.

The initial solution generation uses the distance matrix to create the tours, based on the station loops detected on the layout. The initial solution contains only feasible paths that satisfy the time constraints as well as the capacity constraints of the actual vehicle. To avoid overloaded vehicles, a pre-check method is performed: the vehicle is loaded by all the raw materials in the depot, and then all the other loading and unloading processes are calculated station-by-station to simulate the real milkrun process. In order to calculate with time constraints, both the travel-time (with constant average speed) and the loading/unloading time (time/item) are considered. The purpose of the generated initial paths is to satisfy all the demands by the lowest possible number of feasible cycles. The local search algorithm takes the initial paths and reduces their total time consumption by applying neighbourhood functions where it is possible. The algorithm iterates through all the combinations of the stations, and detects the possible insertions from one path to another. As local search iterates through the paths in a random sequence, different test runs can result different milkrun plans. In order to approximate the global optimum solution, the implemented application performs the local search several times and selects the most appropriate schedule from the generated solutions.

The capabilities of the algorithm were tested in a real production environment. This was a large-scale automotive production system with 2 factory halls, 67 stations and 96 routenodes. The milkrun plan has to satisfy 193 various demands in 60 minutes cycle-time.
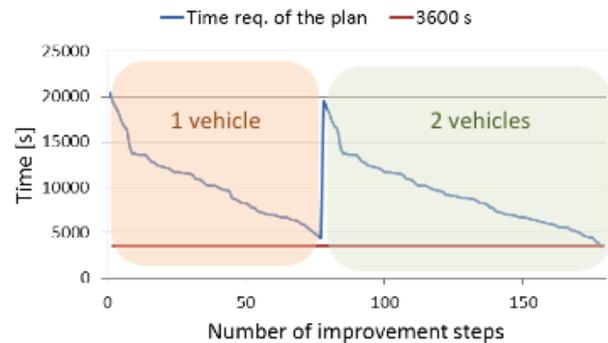
Figure 5: Test results of the case study

The analysed system requires efficient planning process, since the total length of the routes is over 2.5 km that affects critically the total time of the milkrun schedule. The application of inefficient milkrun schedule results low-utilized vehicles and high number of cycles.

In this case, the goal of the test was to minimize the number of vehicles required to perform the services via minimizing the time of the cycles. The proposed initial solution heuristically generated eight feasible paths with 20,400 seconds total time. First, the algorithm tries to optimize the plan by applying one vehicle, and increases the number of the vehicles only if the improvement steps cannot decrease the time requirement of the total plan under the cycle-time (3600 seconds). By this way, the test run could optimize the plan for two vehicles in 177 iterations, and generated a schedule, which time requirement is only 57 minutes *(Fig.5)*.

Conclusions

After an overview about the state-of-the art shop-floor logistics, and the general formulations of the vehicle routing problem, a solution was proposed that uses a novel layout representation and initial solution generation heuristics to solve the milkrun planning problem. In order to demonstrate the capabilities of the

solution, a software prototype was developed and tested on real-life industrial data.

Future work includes the extension of the model with demand partitioning so as to increase capacity utilization of the transportation vehicles. Moreover, handling inhomogeneous demand types (physical aspects of material handling) requires further constraints to be included in the model. Secondly, future work will be dedicated for making the solution available in practical industrial applications. In an on going research project the milkrun planning algorithm will be implemented in a factory- and a logistics planning application currently under implementation.

Acknowledgements

REFERENCES

[1] KOVÁCS A.: Optimizing the storage assignment in a warehouse served by milkrun logistics, International Journal of Production Economics, *2011*, 133, 312–318

[2] LAU H., SIM M., TEO K.: Vehicle routing problem with time windows and a limited number of vehicles, European Journal of Operational Research, *2003*, 148(3) 559–569

[3] DROSTE M., DEUSE J.: A planning approach for in-plant milk run processes to optimize material provision in assembly systems, Proc. 4th CIRP International Conference on Changeable, Agile, Reconfigurable and Virtual Production, *2011*, 605–610

[4] ARNOLD D.: Handbook of logistics, 3rd Edition, Springer, Heidelberg, Germany, *2008* (in German)

[5] WOMACK J.P., JONES D.T.: Lean Thinking, Free Press, *2003*

[6] KLUG F.: Logistics management in automotive industry, Springer, *2010* (in German)

[7] LIN C.: Survey of green vehicle routing problem: past and future trends, Expert Systems with Applications, *2014*, 41, 1118–1138.

[8] BICHLER K., KROHN R.: Gabler kompaktlexikon logistik, 2nd Edition, Springer, Wiesbaden, Germany, *2011* (in German)

[9] SAVELSBERGH M.W.P., SOL M.: The general pickup and delivery problem, Transportation Science, *1995*, 29(1), 17–29

[10] DANTZIG G.B., RAMSER J.H.: The truck dispatching problem, Management Science, *1959*, 6(1), 80–91

[11] PILLAC V.: A review of dynamic vehicle routing problems, European Journal of Operational Research, *2013*, 255(1), 1–11

[12] LU Q., DESSOUKY M.: An exact algorithm for the multiple vehicle pickup and delivery problem, Transportation Science, *2012*, 38, 503–514

[13] BERBEGLIA G.: Feasibility of the pickup and delivery problem with fixed partial routes: a complexity analysis, Transportation Science, *2012*, 46, 359–373

[14] RUSSEL R.: A constraint programming approach to designing a newspaper distribution system, International Journal of Production Economics, *2013*, 145, 32–138

[15] ROSSI F., BEEK VAN P., WALSH T.: Handbook of Constraint Programming, Elsevier, *2006*

[16] CACCHIANI V., HEMMELAMAYR V.C., TRICOIRE F.: A set-covering based heuristic algorithm for the periodic vehicle routing problem, Discrete Applied Mathematics, *2014*, 163, 53–64

[17] VIDAL T., CRAINIC T.G., GENDREAU M., PRINS C.: Heuristics for multi-attribute vehicle routing problems: A survey and synthesis, European Journal of Operational Research, *2013*, 231, 1–21

[18] The VRP Web: http://neo.lcc.uma.es/vrp, (last accessed: May 25, *2014*)

[19] HASHIMOTO H.: Studies on local search-based approaches for vehicle routing problems, PhD Thesis, Kyoto University, *2008*

[20] CARIĆ T.: A modelling and optimization framework for real-world vehicle routing problems, Vehicle Routing Problem, Hrvoje Gold, *2008*

[21] MATAI R., SINGH S., MITTAL M.L.: Traveling salesman problem: an overview of applications, formulations, and solution approaches, Traveling Salesman Problem, Theory and Applications, InTech, *2010*

[22] WILT C.M., THAYER J.T., RUML W.: A Comparison of greedy search algorithms, Proc. 3rd Annual Symposium on Combinatorial Search, *2010*, 129–136

[23] RUSSELL S.J., NORVIG P.: Artificial intelligence: a modern approach, 2nd Edition, Prentice Hall, *2002*

[24] YUE M.: A simple proof of the inequality FFD (L) ≤ 11/9 OPT (L) + 1, ∀L for the FFD bin-packing algorithm, Acta Mathematicae Applicatae Sinica, *1991*, 7(4), 321–331

DESIGN OF A NEW CHEMICAL INJECTION PUMP SYSTEM FOR GAS HYDRATE INHIBITION

Csaba Vörös[1][✉], Viktor Füvesi[1], and Ákos Pintér[2]

[1,2]Research Institute of Applied Earth Sciences, University of Miskolc, Miskolc-Egyetemváros,
POB 2, H-3515 HUNGARY
[1]Department of Research Instrumentation and Informatics, Miskolc-Egyetemváros POB 2, H-3515, HUNGARY
[2]Department of Reservoir and Mining Chemical Engineering, Miskolc-Egyetemváros POB 2, H-3515, HUNGARY
[✉]E-mail: voros@afki.hu

In this study the main feature of chemical injection systems used in gas industry are detailed. The two main energy sources of these systems are air and electric power source. The general structure and main properties of these injection systems are discussed. Examples for a pneumatic commercial and a newly developed, electric system are compared.

Keywords: chemical injection system, hydrate inhibition, solar energy, pneumatic system

Introduction

During the production of gas, a major problem is the formation of hydrate crystals in the pipeline. A considerable amount of hydrate crystal can cause hydrate plugs in the pipeline. The hydrate plug effect lengthens production outages and results in loss of money for the maintainer [1], because elimination of the plug is a time consuming procedure. One of the most widely used traditional solutions to prevent hydrate formation is addition of methanol to the steaming gas. The methanol helps to dehydrate the gas, thus, the growth of hydrate crystal is limited [2]. The methanol can be transported from the gas separator station or gas collection station to the gas wells via a dedicated pipeline. This traditional solution was the most popular practice for a long time in gas industry. This technology is safe, but has several drawbacks, such as the cost of additional pipe to the gas wells, the cost of the methanol regeneration [3], and methanol contamination of the environment. Regardless the methanol technologies are still major solutions for hydrate inhibition for the currently operating gas wells. The modern installations use methanol in low concentrations for newly installed gas wells. These bring the need for injecting methanol locally at the site of the gas wells. Thus, an injection unit is needed for this purpose.

The production-related aspects such as the consumers expecting increased flexibility from gas provider cannot be neglected, because of the habits of the consumers and the appearance of competitive energy-saving technologies. Therefore the different injection systems spread noticeably dynamically in recent years.

In the first part of the paper, the main features of the commercial injection systems are detailed followed by the introduction of a newly developed injection system.

Chemical injection systems

In gas industry there are two applications, where chemical injection is required. The first one is the above described gas hydrate inhibition technology. The second one is where usage of corrosion prevention chemicals is needed. New chemical injection technology appeared in sophisticated form using a wide service range as a result of a change in gas production and distribution in the last decade.

Structure of pneumatic chemical injection pump systems

The most commonly used chemical injection units in gas industry employ the energy of gas for dosing. These systems are called "operating without auxiliary power systems". In details, the energy of gas steam in the pipeline is used as a power source in the dosing of hydrate inhibiting chemical pump systems. This method has the major advantage that the gas is always available, or it is not necessary to provide a separate power source for the equipment to operate.

Big disadvantage of the technique is that gas is required for operation and this gas is emitted to environment, which is a harmful effect, moreover, the cost of it is also not negligible. There are some solutions to utilize the emitted gases, but these methods raise the price of the pump.
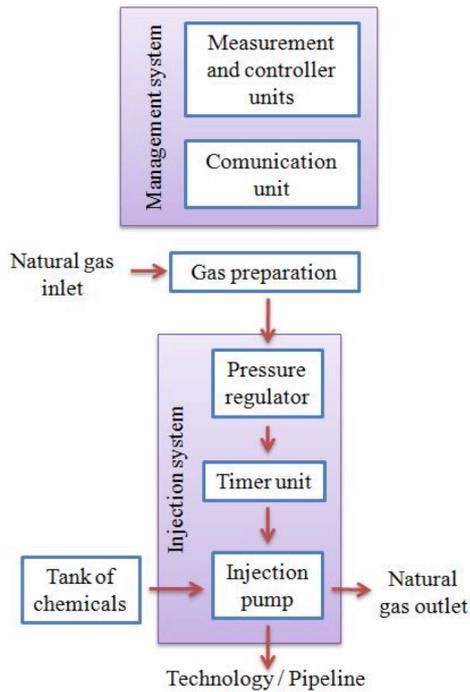
Figure 1: Block scheme of chemical injection system
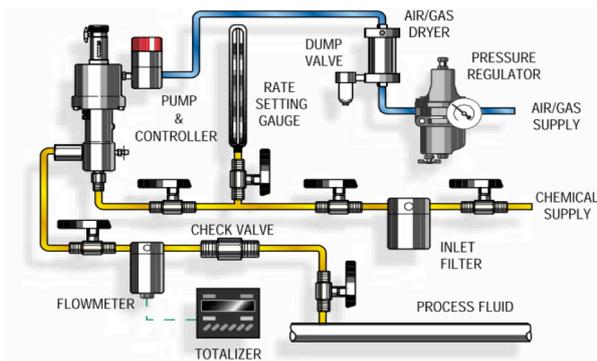
Figure 3: Pneumatic injection system from HASKEL MILTON ROY [4]

Fig.1 shows a common structure of chemical injection pump systems, which illustrates how the actuated gas gets into the injection system through a gas preparation unit. This unit is typically a special tank, which is filled with methanol. By using this type of conditioning one can be sure that the gas is sufficiently dry, thus, the water does not disturb the operation of the pump. It is needed to avoid the risk of freezing out various components. The gas bubbles through the methanol, so it loses a significant part its water content. The replacement of the methanol is needed at appropriate intervals of time as a technology management. After the gas leaves the conditioning unit, the gas steams through a two-stage pressure regulator. This unit provides the required pressure to the system for operation. The timer unit is responsible for providing appropriate technological demands of the schedule, i.e. sufficient number of strokes is needed to operate the pump. The chemical is injected with the piston of the injection pump. The injection system includes other components, which are essential for the
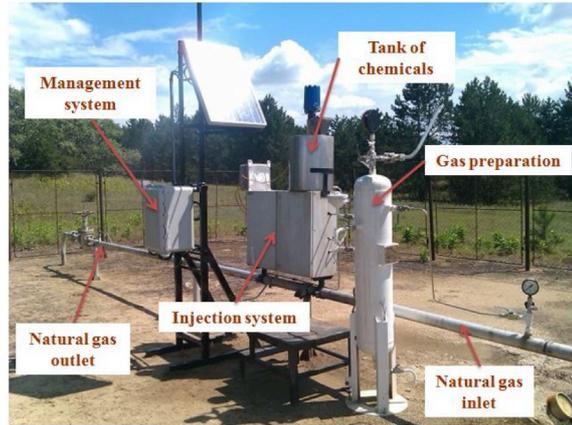
Figure 2: Chemical injection system on a gas well

operation, such as check valves, ball valves and other fittings. The hydrate inhibitor compound can be found in the tank.

The main task of the control system is the monitoring of the injection system, so several parameters are required to be measured, e.g. the level of chemical in the tank, the pressure of the after the regulator and the ambient temperature. One of the most substantial information about the injection pump is the entry of the inhibitor to the pipeline. To measure these parameters, transducers are needed, which require some solar energy for the operation. *Fig.2* shows a well-organized, modern gas well in Hungary. The injection system used employs the gas of the well for its operation.

Injection system from HASKEL MILTON ROY

The HASKEL MILTON ROY company develops and manufactures various injection systems for special oil and gas applications for many years. Chemical Injection Pump (CIP) series is one of the most widely used equipment for gas wells. A pneumatic version of the device is shown in *Fig.3*, as a typical system. The driving gas or air is fed from the regulator to the gas-drying unit. The air is let into the injection pump, which includes the timer unit. The chemical enters the pump through the appropriate valves and filters. The dosage of chemicals can be adjusted through the rate setting gauge unit (typically it is a burette) at, which the quantity of chemical of one injection is displayed. The strokes can be counted by using the flow meter on the system. The check valve prevents the injection pump from the backpressure of the process fluid. The company can deliver the complete unit in different versions according to the quantity of injection, e.g. the given series can be adjusted from 0.27–25.8 dm^3 min^{-1}. The range of operation pressure is also wide, typically 120–600 bar.

Structure of electric injection systems

This category includes systems, where the electricity is the only energy source. In general, the site of the gas
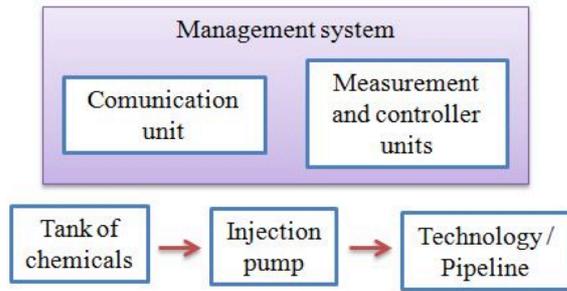
Figure 4: Block scheme of electric power injection system
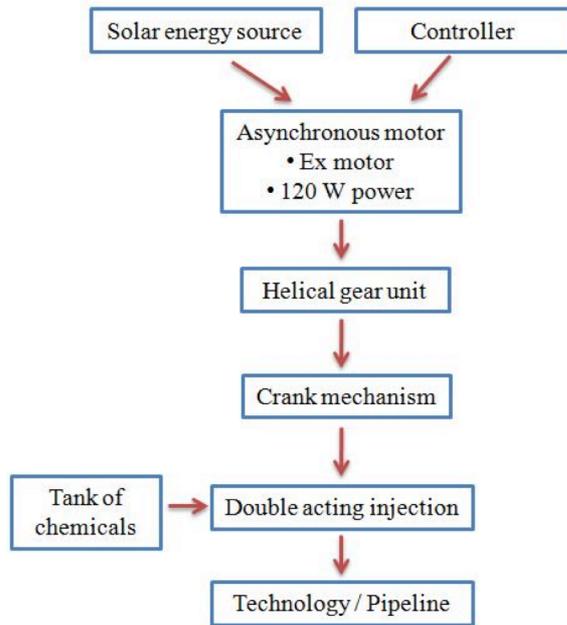
Figure 5: Block scheme of new injection system

well is not equipped with electric power because of the relatively large distance from the processor (separator station, collector, etc.). Thus, these systems are not commonly employed in the gas well applications.

Most of the electric chemical injection pump systems are based on the principle of displacement pumps. These high-pressure pumps provide the chemical injection by using speed adjustment in a wide range, or with different piston volumes.

As shown in *Fig.4*, the electrical injection pump system contains relatively few units. It cannot be ignored that the electric pump does not only include a motor-driven piston with frequency inverter, the controller electronics is also an important part of the system. Solar energy can be applied as energy source of this type of system. This source can feed the control electronic besides the main actuator of the system.

In-house designed injection system

The development of the injection system was a part of a bigger project with the Scada Ltd, a company from automation industry. The aim of development was to

Table 1: Main parameters of the injection pump designed

Asynchronous motor	
Nominal voltage	230 V(Y) /400 V (D), 50 Hz
Nominal power	0,12 kW
Ex class	PTB 07 ATEX 1058 X II2G Ex de IIC T4
Type	DEx 63K/4K
Vendor	HEW
Helical gear unit	
Gear ratio	41,58
Ex class	Ex II 2G c IIC T4 X
Type	SK02F-IEC63 /26-63 S/4 TF/2G
Vendor	Nord
Crank mechanism	
Length of stroke	20 mm
Ratio of mechanics	7
Injection unit	
Diameter of piston	5 mm
Vendor of seals	Trelleborg
Type of seals	Turcon® Variseal®

create an injection system, which can be used mainly at Hungarian gas wells. Thus, the temperature requirement of the system was in the -40 °C to 60 °C range. The system must be capable of working in 'Ex' environment with high efficiency. The power source of the actuator is solar energy to reach close to zero emission of the system. In the system, corrosive chemical has to be injected, thus, a corrosive resistant is required at critical parts of the equipment e.g. piston, piston space, etc. The maximal working pressure was 160 bars. A wide range of injection volume was also a fundamental requirement. These requirements were the main parameters at the design phase of the equipment. The block scheme of the developed system can be seen in *Fig.5*.

The base of the system is an Ex class asynchronous motor *(Table 1)*, which is assembled with a helical gear unit from Nord manufacturer. The base of injection system is a crank mechanism, which transforms the rotary motion to alternating motion. The pump has two pistons to reach higher injection volume and a well-balanced load of the motor. The main parameters of the mechanics can be found in *Table 1*.

The sealing of the mechanism is a critical design point. The relatively high pressure, corrosive fluids, and high stroke number require special solution for sealing. A single acting, spring energized-seal was selected. The application conditions demanded the tight guiding of neck of piston, to avoid the damage of seals during the operation of pump *(Fig.6)*. The electromechanical actuator gives the opportunity to smoothly alter the injected volume in continuous or in periodic operation. *Table 2* shows the injected volume of the pump at different frequency of the inverter in continuous mode. The power source of the system is an array of solar cells. The power of solar energy system was 2 kW to get extended service and short recharge time. The capacity of the accumulators was 800 Ah. The developed equipment has a PLC-based control system to monitor the main parameters of the technology and to change the behaviour of the injection system.
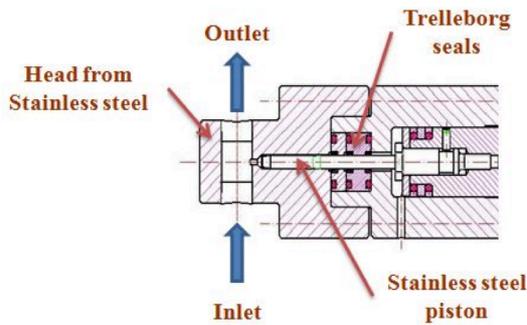
Figure 6: Section of injection unit

Table 2: Injected volume of pump at 150 bars

Frequency of inverter, Hz	Injected volume, $dm^3\ h^{-1}$
30	21.71
50	36.19
70	50.67
90	65.14

The measured parameters are the pressure of siphon of well, drill pipe, injection pipe; the temperature of pipe of well, soil, chemicals, controller, temperature of inverter; the level of the chemical in the tank; the current and the voltage of the inverter. The frequency of the inverter can be changed in function of the conditions of the well.

The injection system is able to communicate wirelessly with a server. By using modern computer technology, the main parameters of the system can be monitored and changed. *Fig.7* shows the website of the system, which shows, on one hand, the main actual online parameters of the system, on other hand, the trends and history of the parameters on diagrams. The structure of the website is user friendly, simple. A model technology for an actual gas well was set up at the site of the Scada Ltd. near to Hajdúszoboszló *(Fig.8)*. At the test site, main parameters like level of the chemical in the tank, process gas pressure, frequency of the inverter, etc. were monitored and measured. After the successful test of the model technology, the equipment was moved to a real gas well at Békásmegyer in October 2012.

Conclusion

A new solar energy-based chemical injection pump system was developed with a modern control and computerized support system at University of Miskolc, Research Institute of Applied Earth Sciences. The equipment meets the modern requirements. It is an efficient injection system; it has wide operation range and has user-friendly high-level computer support. In the process automation field of gas industry, the adaptation of new approaches appear slower in consumer or in industrial environments than in other fields of industry. The new, energy efficient, solar supplied, standalone chemical injection systems hopefully can enter in the field and will be utilized more wells in Hungary.
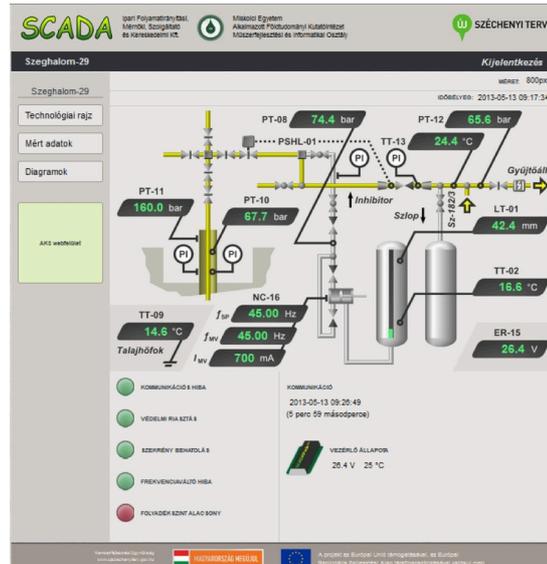
Figure 7: Website of pump system
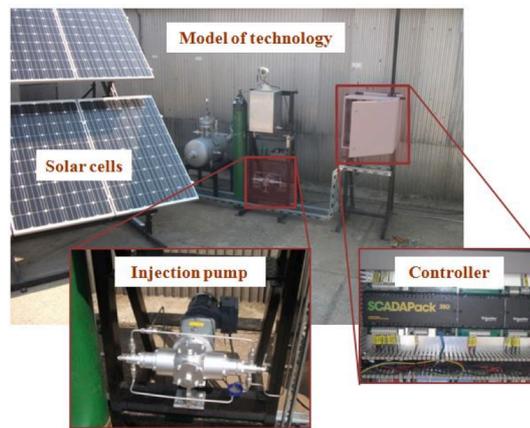
Figure 8: Block scheme of chemical injection system

Acknowledgement

REFERENCES

[1] BERECZ E., BALLA-ACHS M.: Gázhidrátok, Akadémiai Kiadó, Budapest, Hungary, *1980* (in Hungarian)

[2] HAMMERSCHMIDT E.G.: Formation of gas hydrates in natural gas transmission lines, Ind. Eng. Chem. *1934,* 26(8), 851

[3] LEDERHOS J.P., LONG J.P., SUM A., CHRISTIANSEN R.L., SLOAN, E.D. JR.: Effective Kinetic Inhibitors for Natural Gas Hydrates, Chemical Engineering Science, *1996*, 51(8), 1221–1229

[4] Williams Chemical Metering Pumps: www.willamspump.com (last accessed: May 25, 2014)

IONIC LIQUIDS IN ADVANCED ENERGY STORAGE CELLS

ATTILA GÖLLEI ✉ AND ATTILA MAGYAR

Department of Electrical Engineering and Information Technology, University of Pannonia, Egyetem Str. 10,
Veszprém, H-8200, HUNGARY
✉E-mail:gollei.attila@virt.uni-pannon.hu

The interest in the research and application of ionic liquids has been increased worldwide, especially in petrochemical industry, nuclear industry, oil and gas industry, chemical, and electro-chemical industry. In the field of chemical operations in microwave field with ionic liquids (organic chemical synthesis, catalytic operations, etc.) new horizons opened that were supported by the dramatically increasing number of studies dealing with this question and their impact. Due to the high durability of ionic liquids as a function of temperature, a wider temperature range of operations can be accessed. In addition, it offers environmental friendly solution in the replacement of the toxic solvents with generally low volatility. The area of application extends to electrolytes of novel battery cells. The important prerequisite for application of ionic liquids is the knowledge of their physical and chemical properties. Our aim was to test the dielectric properties, viscosity, and temperature dependence of the electrical conductivity. Considering our measurements, it can be declared that the aforementioned properties show significant temperature dependence in the case of ionic liquids. This knowledge is important for the usability, design and execution of production and related optimization processes. It is particularly important that energy storage cells can be exposed to large temperature fluctuations. This study discusses the sample materials, the usage possibilities of the sample materials and the obtained results.

Keywords: Ionic liquids, energy storage cells, conductivity, temperature dependence, viscosity

Introduction

The key parameters in the modelling of microwave behaviour of ionic liquids are the dielectric constant, dielectric dissipation factors, in addition the parameters characterizing the polarizability and the microwave energy absorption. Traditional measurement methods of the static dielectric constant of ionic liquids fail, because they cause intense shortcut due to their high electrical conductivity. According to literature, these contexts have not been studied systematically because of the novelty of this subject. The significance of these results is critically important in the planning and controlling chemical reactions.

Ionic liquids (IL) opened new technological possibilities at the end of the 1990s providing a wide range of applications, especially for areas of chemistry and particularly in green chemistry. ILs are beneficial due to their special chemical and physical characteristics. Nowadays, the use of energy storage units, particularly mobile phones, electric vehicles, and uninterruptible power supply systems has more and more importance. Recently, room temperature ionic liquids (RTILs) have been extensively studied as electrolytes of lithium ion batteries from cellular phones to electric vehicles. The unique properties make RTILs attractive candidates for the electrolyte base of safe lithium battery [1, 2].

Generally, RTILs are well known to be thermally stable and non-flammable and their nature can hopefully improve the safety of electrochemical devices with aprotic solvents, such as Li batteries and super-capacitors.

1-Ethyl-3-methyl-imidazolium tetrafluoroborate (EMIM-BF$_4$) has a high ionic conductivity, which is comparable to those of organic solvent electrolytes. In addition, the viscosity of EMIM-BF$_4$ is low. Therefore, EMIM-BF$_4$ is expected to be a good electrolyte candidate for Li batteries. We have found that a Li/LiCoO$_2$ cell with RTILs as an electrolyte base works reversibly, indicating that QAimide RTILs are quite stable even at the Li reduction potential. Clearly, RTILs improve the safety of Li-ion batteries with carbon anodes. To date, RTILs are the most promising electrolytes for improving the safety of Li-metal batteries, which could be important for higher energy densities [3].

A set of four imidazolium ionic liquids (solid at room temperature) and one imidazolium ionic solid were screened for their potentials as electrolytes in manganese dioxide free LECLANCHÉ batteries, equipped with a zinc anode and graphite cathode [4].

It should be noted that the ionic liquid 1-ethyl-3-methylimidazolium hydrogensulphate (EMIM-HSO$_4$) similar to many other ionic liquids, such as those based on nitrate and dihydrogenphosphate anions, have also been found they work well in this battery design. Novel
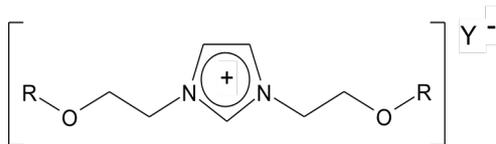
Figure 1: The structure of investigated ionic liquids

batteries are designed using standard cathode materials such as MnO_2, PbO_2, NiO and AgO, and anode materials such as Zn, Sn and Pb. Additionally, by using a solid polymer electrolyte composed of polyvinyl alcohol and anionic liquid, new types of solid state batteries were demonstrated with discharge voltages ranging up to 1.8 V, depending upon the type of cathode and anode used [5].

Ionic liquids like 1-butyl-3-methylimidazolium tetrafluoroborate (IMIM-BF4) or hexafluorophosphate (IMIM-PF6) and 1-butyl-4-methylpyridinium tetrafluoroborate (PyBF4) were mixed with organic solvents such as butyrolactone (BL) and acetonitrile (ACN). The specific ionic liquid used in the given study is shown in *Fig.1*. A lithium salt ($LiBF_4$ or $LiPF_6$) was added to these mixtures for possible application in the field of energy storage (batteries or super-capacitors). Viscosities, conductivities and electrochemical windows at a Pt electrode of these electrolytes were investigated. All the studied electrolytes are stable toward oxidation and exhibit a vitreous phase transition, which has been determined by application of the conductivity measurements. Mixtures containing the BF_4^- anion exhibit the lowest viscosity and highest conductivity [6].

Methods

The essential precondition for the application of ionic liquids in batteries is a preliminary examination of its properties. Therefore, every measurement method available to us was used on a number of ionic liquids examined. We have carried out microwave measurements, such as temperature change, dielectric properties at 2.45 GHz frequency, electrical conductivity according to temperature and viscosity changes depending on the temperature.

The velocity of temperature change of ionic liquids was measured in the CEM Discover unit. This is a well-known and commercially available apparatus in microwave chemistry. The device has a cylindrical operating space and on the cylindrical peripheral surface there are many slots where microwave energy can enter. This solution provides a special high homogeneity of the microwave field. The temperature was measured at the bottom of the compartment with an infrared thermometer. The amount of the samples used was 0.5 g, which was placed in an inner diameter of 12.5 mm of borosilicate cylindrical glass flask.

Table 1: The formulas of investigated ionic liquids

Compound	R	Y
1	CH_3	BF_4
2	C_2H_5	BF_4
3	C_3H_7	BF_4
4	C_4H_9	BF_4
5	CH_3	PF_6
6	C_2H_5	PF_6
7	C_3H_7	PF_6
8	C_4H_9	PF_6
9	C_2H_5	CI
10	C_2H_5	Br
11	C_2H_5	SCN
12	C_2H_5	$N(CN)_2$
13	C_2H_5	$N(SO_2CF_3)_2$

It is important to know that during microwave treatment the microwave energy interacts with the treated material, and the material converts the electric energy into thermal energy according to its characteristic of dielectric properties. During this conversion process, the macroscopically observable result is the measurable increase of the temperature in the treated material. The rate of the temperature increase depends on the microwave field and the treated material properties, which is described in the following equation:

$$\Delta T / \Delta t = P_v / \rho C_p = iE^2 f \varepsilon'' / \rho C_p \qquad (1)$$

where, the $\Delta T/\Delta t$ is the velocity of temperature change in Kelvin per second, P_v is the absorbed power in the sample, ρ is the density of the sample, C_p is the specific heat of the sample, E^2 is the strength of the electromagnetic field in the sample, f is the frequency of the field and ε'' is the dielectric loss, i means this is a complex value [4]. *Eq.(1)* shows that the temperature of the treated material is influenced by a number of factors. The separate testing of their effects is not an easy task, since ρ, ε, C_p features themselves are temperature dependent, and this value is difficult to accurately measure inside a material. In case of strictly homogeneous series of examined compounds the situation is more simplified, because some simplifications are allowed.

Assuming ρ and C_p does not change significantly in the function of temperature, the multiplication ρC_p was almost considered to be constant, so that the rate of temperature change is determined by E and ε'' only. Further simplifications can be made if the device is single-mode with the same volume and shape, and the microwave energy is constant during the investigation. The value of E is determined by ε' and ε' and ε'' and the rate of temperature change is essentially determined by ε'' value.
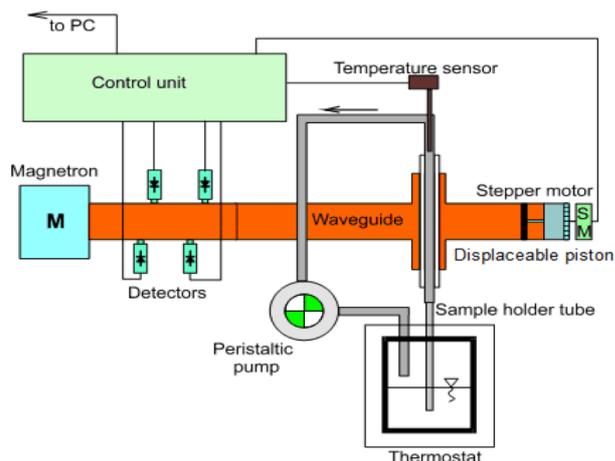
Figure 2: Scheme of the experimental setup

Figure 3: Temperature profile for 1-butyl-3-methylimidazolium tetrafluoroborate at 3W

Measurement setup

The schematic representation of the in-house designed experimental set-up for automated online measurement of dielectric properties of ionic liquids in a definite temperature range is shown in *Fig.2*. It is composed of the following devices and instruments; cylindrical sample holder unit, thermostat, peristaltic pump, waveguide, temperature sensor, displaceable piston, stepper motor, magnetron, detectors, control unit, and a PC.

The IL sample, which is placed in the thermostat to be kept at the desired temperature, is passed across the waveguide having a length of about 3λ through the sample holder tube with the help of a peristaltic pump. The electric energy is transformed into microwave energy by the magnetron. The stepper motor is controlled by a microprocessor control unit equipped with an Intel 8-bit microcontroller, a 12 bit A/D converters for receiving the four diode-detector signals, a stepping motor driver, and a RS-232 serial interface to connect it to a PC. The control unit collects the detector signals and the temperature data from the temperature sensor and sends them to the PC. Furthermore, it controls the position of the short circuit displaceable piston on the basis of algorithm software elaborated for this purpose.

The measurement method is based on the compensation of phase change due to the microwave energy absorption of the liquid sample. The short circuit piston situated behind the sample must be actuated for compensation. The energy conditions created by the wave front in the waveguide are measured by four diode-detectors.

Dielectric constants, dielectric loss factors and the temperature dependence of the dielectric properties of ionic liquids intended to be used in batteries were determined by the above described self-designed microwave dielectrometric apparatus *(Fig.3)* at the frequency of 2.45 GHz and at different temperatures (30°C, 40°C, 50°C, 60°C, 70°C, 80°C, 90°C, 100°C, 110°C and 120°C). The speed of the change in temperature depends on the electrical field strength in the material *(E)*, the absorbed microwave power, *(P_v)*
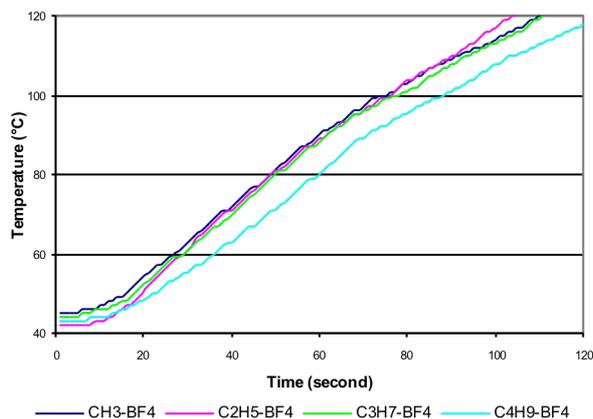
density *(ρ)*, the specific heat capacity *(C_p)* and the dielectric loss factor ε'' and can be given by *Eq.(1)* [7].

Results and Discussions

The effects of R groups *(Table 1)* were examined in case of two anions - BF_4 **(1–4)** and PF_6 **(5–8)** at 3 to 5 W of microwave energy. The results indicate that the decrease of the number of C atoms in R increases the rate of temperature change. By increasing the microwave power significantly the temperature rapidly converge and sometimes crossing over, but this is not typical in *Fig.3*. Overall, the rate of temperature rising follows the order of ε' and ε'' suggesting that in a strictly homologous series the temperature change of ionic liquids is determined by ε'' [5].

After evaluating the rate of temperature change, we investigated two additional properties of the ionic liquids used in the advanced batteries. These parameters were the viscosity and the electrical conductivity. Both of these parameters are the functions of temperature. They were examined between 20 °C and 100–120 °C temperature range. The viscosity was measured up to 95 °C due to the sample material being heated by a water-filled heat exchanger. The measurements were performed in a SV-10 type of vibration viscometer (A&D Ltd. Japan). *Fig.4* shows that the viscosity at room temperature is high, typically several hundred Pa·s. It decreases rapidly with the temperature increase similar to a reciprocal curve. At 100 °C, the value of viscosity can be ten times lower then at the initial value as can shown in *Fig.4*.

The investigation of the viscosity is important, since it basically correlates with the mobility of ions in the electrolyte, which can affect the conductivity, and thus the internal resistance of the battery cell. By cooling ionic liquids, the viscosity is increasing rapidly. Around freezing temperature the ionic liquids were already similar to a "honey density" mass.

Since the energy storage cells are used in a wide temperature range, it is important to know that the
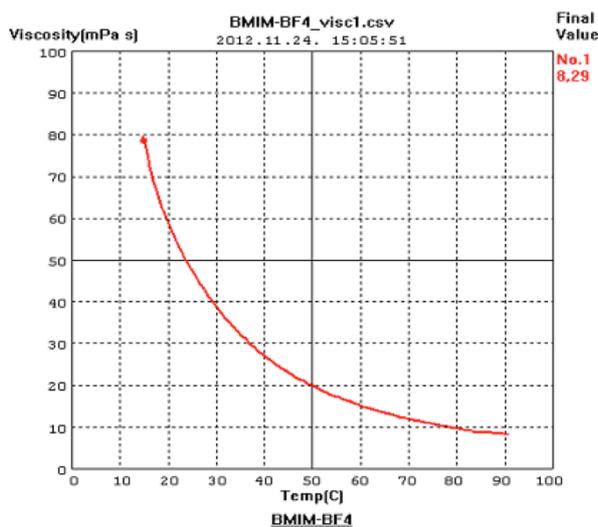
Figure 4: Viscosity of BMIM-BF4 as a function of the temperature
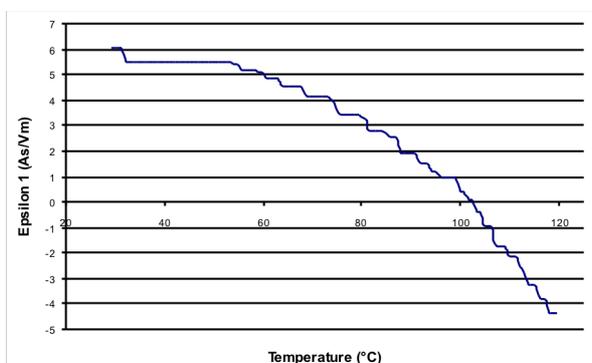
Figure 6: Dielectric constant of BMIM-BF4 as a function of the temperature

Figure 5: Conductivity of five compounds as a function of temperature
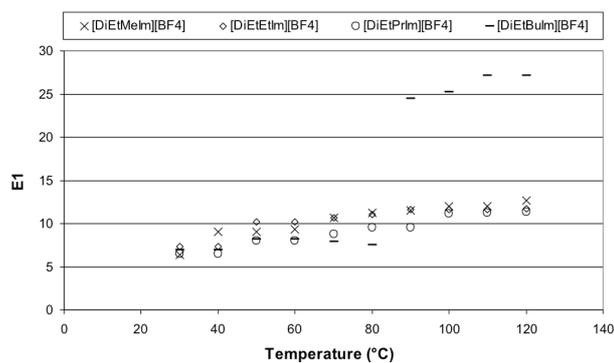
Figure 7: Dielectric constant of ILs containing common BF$_4^-$ anion and different cations

electrical conductivity of the electrolytes is a function of temperature.

Fig.5 contains five different electrical conductivity plots of ionic liquids in the function of temperature. It can be seen that the above mentioned BMIM-BF$_4$ does not have the highest conductivity value, but the depending on the temperature the conductivity value of this material increases with the greatest intensity [6].

A new and surprising result can be seen in *Fig.6*. The dielectric constant value of BMIM-BF4 reaches a value of 1 at close to 100 °C, similarly to the value of the vacuum and with the further raising of the temperature this value decreases into the negative range. This fact suggests that by increasing the temperature the electrical conductivity is rising greatly and this compound is no longer an insulating material, but a conductor. Therefore it is important to keep the temperature in an adequate range when batteries are used and especially in the charging period.

The dielectric constant *(E$_1$)*, dielectric loss factor values of ILs built up of the same BF$_4^-$ anion and six previously described cations containing alkyl chains with different lengths (DiEtMeIm$^+$, DiEtEtIm$^+$, DiEtPrIm$^+$, and DiEtBuIm$^+$) at different temperatures between 30°C and 120°C are shown in *Fig.7*.

At the initial measuring temperature of 30 °C all four of the studied ILs have similar dielectric constants around 7. With the increase in temperature the E_1 values for DiEtMeIm(BF$_4$), DiEtEtIm(BF$_4$) and DiEtPrIm(BF$_4$) slightly increase up to 13, while the dielectric constant for DiEtBuIm(BF$_4$) shows a sudden break at 90°C and at 120°C it reaches the value as high as 27. This could be explained with some sudden changes in the structure of the ILs or in the physico-chemical interactions between the anion and the cation. Excluding the results for DiEtBuIm(BF$_4$) the values at elevated temperature show that the highest E_1 value belongs to DiEtMeIm(BF$_4$), followed by DiEtEtIm(BF$_4$) and then by the DiEtPrIm(BF$_4$), hence the dielectric constant increases with the decrease in the alkyl chain length of the cation.

Taking into account the dielectric constant, dielectric loss and electrical conductivity results of all the investigated ILs it can be concluded that the alkyl chain length of the cation and the structure of the anion strongly influence the dielectric properties of the ILs and that the highest *G* value is exhibited by the DiEtMeIm(BF$_4$) IL at 30°C temperature, hence it is the most suitable candidate for battery applications. Furthermore it can be stated that implementing the knowledge about the connection between the IL structure and the dielectric properties another ILs should be studied in order to select the ones most adequate for electrolyte application.

Conclusion

Due to the special characteristics of ILs, such as wide electrochemical windows, high inherent conductivities, high thermal and electrochemical stability, tuneable physicochemical properties, etc., they are potentially excellent candidates for environmentally sound, green electrolytes in batteries. In order to predict their success in a specific application, it is essential to gain information about their dielectric properties.

However, several techniques have already been published about determination of the dielectric properties of ILs, so far no experimental methods were introduced due to the fact that such measurements cannot be realized by classical methods because of the high conductivity of ILs.

Summarizing our results, we can conclude that when ionic liquids in the energy storage cells were used, their physical, chemical and electrical parameters are significantly temperature dependent. In any case, when they are applied, we should specify the range of application temperature. Stepping out of this range of the energy storage cell may not meet the expected specification values, or in the worst case it may be permanently damaged.

Acknowledgements

REFERENCES

[1] ALARCO P.J., YASER A.L., RAVET N., ARMAND M.: Lithium conducting pyrazoliumimides plastic crystals: a new solid state electrolyte matrix. Solid State Ionics, *2004*, 172(1–4), 53–56

[2] SCHIFFMANN R.F.: Microwave and dielectric drying, in: 2nd Handbook of Industrial Drying, (Ed. MUJAMDER A.S.), Vol 1, Marcel Dekker, New York, USA, *1995*, pp. 345–372.

[3] HIKARI S., HAJIME M., KUNIAKI T.: Application of room temperature ionic liquids to Li batteries Electrochimica Acta, *2007*, 53, 1048–1054

[4] ZHANG Z., GAO X., YANG L.: Electrochemical properties of room temperature ionic liquids incorporating BF_4 and TFSI anions as green electrolytes, Chinese Science Bulletin *2005*, 50(18), 2009

[5] GAO X.H., ZHU M.J., ZHANG Z.X.: Research on room temperature molten salts with alkylated imidazolium salt as electrolytes, Chemical World (Supplement, I), *2004*, 45, 156–157 (in Chinese)

[6] NISHIDA T., TASHIRO Y., YAMAMOTO M.: Physical and electrochemical properties of 1-alkyl-3-methylimidazolium tetrafluoroborate for electrolyte, J. Fluor. Chem., *2003*, 120, 135–141

[7] GÖLLEI A., VASS A., PALLAI E., GERZSON M., LUDÁNYI L., MINK J.: Apparatus and method to measure dielectric properties (ε' and ε") of ionic liquids. Review of Scientific Instruments, *2009*, 80(4), 6734–6748

Advertise upcoming meetings,

conferences, and workshops;

make public announcements;

introduce your research laboratory;

a new product or a service

in the

Hungarian Journal of Industry and Chemistry

Please contact us if interested!

MONITORING AND DIAGNOSIS OF MANUFACTURING SYSTEMS USING TIMED COLOURED PETRI NETS

ADRIEN LEITOLD[1], BRIGITTA MÁRCZI[2], ANNA IBOLYA PÓZNA[2], AND MIKLÓS GERZSON[2✉]

[1] Department of Mathematics, University of Pannonia, Egyetem str. 10, HUNGARY
[2] Department of Electrical Engineering and Information Systems, University of Pannonia, Egyetem str. 10, HUNGARY
✉E-mail: gerzson@almos.uni-pannon.hu

Novel fault modelling and integration method were applied in the case when the faultless operation of the system was modelled by a high-level, coloured Petri net. In order to achieve realistic investigations, a timed coloured Petri net model of the system was constructed, where faults can occur in the manufacturing lines. The faultless and fault containing models were implemented in CPNTools both for non-timed and timed cases. The resulted model was investigated both via simulation and using the occurrence graph. For efficient analysis of the occurrence graph a software module called OGAnalyser was developed.

Keywords: monitoring and diagnostics of manufacturing processes, coloured timed Petri nets, probabilistic models, occurrence graph

Introduction

Models are often used for the description and investigation of complex systems even if they cannot perfectly describe the investigated system. The course of a manufacturing system can be split up to distinct steps of serial or parallel technological sub-processes. This enables the description of the system by a discrete event systems model [1] in the form of Petri nets. During the design process of a manufacturing system, often only a model of the faultless operation is created. The integration of possible faults into the model could give important information for more complex investigation of the system. In our previous work [2] we have integrated fault events with different occurring possibilities into low-level Petri net models of manufacturing systems in such a way that the size of the model remained almost the same.

In our recent work, we applied the above fault integration method for the case when the faultless operation of the system was modelled by a high-level, coloured Petri net (abbreviated as CP-nets) [3]. In low-level Petri nets the transitions fire instantaneously, but the events of a real system take place for a certain amount of time influencing the operation of the system. Therefore, a timed coloured Petri net model of the system is constructed in order to achieve realistic investigations.

CPNTools [4] offer tools for modelling and analysing of CP-nets. There are two possibilities for the investigation of a manufacturing system in CPNTools: the simulation and the analysis of the occurrence graph. In case of fault modelling using different occurring possibilities; however, the standard occurrence graph does not give information about the probability of the different occurring states of the system. Therefore, a special software module, called the OGAnalyzer has been developed for solving this problem. Weights can be assigned to the arcs of the occurrence graph and the software calculates the probability of each node in the occurrence graph, i.e. of each operational state of the system.

Petri net model of a manufacturing process

Petri nets enable both the mathematical and the graph representation of a discrete event system to be modelled, where the signals of the system have discrete range space and time is also discrete [5]. Petri nets can be used for describing a controlled or open loop system for modelling the events occurring in it, and for analysing the resulted model. For the different application purposes, various modifications of the original Petri net were developed with the aim of improving the modelling capabilities. One of the approaches is the *coloured Petri nets* (CP-nets). We use here the CP-nets for modelling technological systems and their diagnosis, i.e. for the determination of faulty operational modes of the investigated system.

Coloured Petri nets

The CP-nets combine the modelling advantages of Petri nets and the compactness of the functional

programming language Standard ML [6]. A Petri net is bipartite graph having circles and rectangles as nodes. Circles refer to the 'places' in the net and rectangles to the 'transitions'. Places represent the state of the elements in the modelled system, while transitions correspond to the actions taken place in it. There are 'arcs' between places and transitions referring to logical relations of the system. If an arc directs from a place to a transition then the place acts as a precondition of the given transition while the arcs in the opposite direction represent consequences of transitions. Each place can be marked with one or more coloured tokens representing the state of the modelled element.

Here we emphasize two important novelties of CP-nets only. The tokens describing the state of the system have data value, the so-called *token colour* attached to them. In this paper these colours are used to identify workpieces and to describe the operation to be performed on them. Places, transitions and arcs can have 'inscriptions'. An inscription of a place determines the set of colours that a token on the place can have. Another place inscription gives the actual number of the tokens on the place, i.e. the current marking of that place. The inscriptions of transitions can contain different types of functions. These functions determine the type of the colour set of the incoming and outgoing tokens and the operation performed on them. The arc inscriptions can be used for evaluating the result of the performed action at the previous transition. These conditional expressions define the colour of the token on the following place.

The original Petri net concept did not contain the time; however, the firing of transition takes place instantaneously. In case of real technological systems the time has a great role during the occurring of events. CP-nets also offer the possibility of adding time to the operation of transitions. The firing rule of a transition in *timed Petri nets* is as follows: a transition is enabled if all of its input places contain sufficient number coloured tokens defined by the arc functions and the allotted time has elapsed.

One of the main advantages of modelling with Petri nets is the ability of describing sequences of discrete events that occur both in a serial and in a parallel way. In case of parallelism, we can distinguish two different situations. In the first case the two or more series of events can take place independently of each other. This situation occurs when workpieces can be elaborated in different manufacturing lines in parallel way. In the other case only one of the event sequences can take place because these events exclude mutually each other. These events have the same precondition, and the occurrence of any of them makes this precondition invalid. This kind of parallelism is called 'conflict situation'. In a Petri net the conflict can be recognized when a place is the precondition of two or more transitions. In this case it is randomly selected, which transition takes place. The conflict can have two different sources in technological systems. A conflict occurs in a technological sense when two or more processes want to use the same tool or resource e.g. a

robot. Usually it is worth to assign priority to each conflicting transition in order to define their sequence.

When a fault occurs during the operation of the system, it also causes a conflict situation. This can be avoided by adding a special probability function to arc expression functions. By evaluating this function, the occurrence of the fault can be unambiguously determined during execution.

Analysis of Petri nets

There are two basic directions for the analysis of a Petri net: (i) the structural method, which is independent of the initial state of the net and (ii) the investigations based on a given initial state (the behavioural analysis). In this paper the latter is used for the investigation of technological system behaviour.

Simulation is our primary tool for the checking the correctness of a model. Starting from a given initial state the user can check whether the operation of the system terminates in the appropriate state. It can also be investigated, which transitions become enabled in certain steps, whether there is a conflict among then. Simulation investigations do not give unambiguous answers to questions of formal analysis formulating in Petri net literature [7] but they complement them well.

Another Petri net analysis method uses the 'occurrence graph'. The basic idea of the occurrence graph is to construct a graph, which contains all of the reachable markings from a given initial state. These marking are the nodes of the occurrence graph and the arcs connecting the nodes refer to the logical relations realized by the firing transition between two markings. Unfortunately, the occurrence graph even of a small Petri net may become very large. Therefore, several reduction methods were proposed in order to get a relatively small occurrence graph [7]. Most of the simulation tools, as the CPNTools [4] used by us, are able to construct the occurrence graph.

Modelling and analysis of technological systems

In the following, we wish to demonstrate the use of CP-nets and their occurrence graphs for modelling and analysis of technological processes. Both for normal faultless and faulty mode operation of the technological system are considered in the non-timed and timed cases. The analysis is performed via simulation and with different investigations of the occurrence graph.

The manufacturing system and its operating procedure

A simple case study is presented here for a manufacturing system containing two manufacturing lines and a robot. The arrangement of the system can be seen in *Fig.1*.
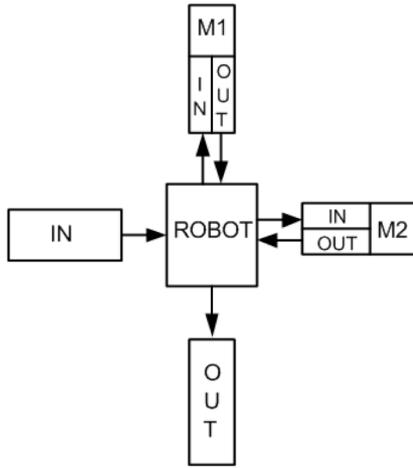
Figure 1: The manufacturing system

The workpieces to be processed appear on the input place IN. The task of the robot is to put them to the appropriate input place of a manufacturing line `M1_IN` or `M2_IN` according to operational instructions. Assume that the two manufacturing lines perform different actions on the workpiece of interest. When the manufacturing process is over, the finished workpiece appears at the output end of the line either on `M1_OUT` or `M2_OUT` depending on the performed action. If the workpiece has to be modified on the other manufacturing line then the robot puts it onto the other input place. If the manufacturing process is over then the robot puts it into the product container OUT. Assume that one workpiece at a time can be on the input and output places of manufacturing lines. It follows that the

robot can only transfer a workpiece from place IN to the input place of a manufacturing line if this place is empty and the precondition of the start of a manufacturing process is that the output place of this line should be empty. As a general precondition of all transfer processes the robot has to be free.

The CP-net model of the normal (faultless) operation of the manufacturing system in the form of a screenshot from CPNTools can be seen in *Fig.2*. The process starts with a token at place `START`. The transition `generator` generates the prescribed number of tokens representing workpieces at place `IN`. The colour assigned to a token contains an identifier of the workpiece and a code referring to the manufacturing process or processes to be carried out. Four kinds of manufacturing mode are possible in this manufacturing system: the workpiece has to be processed on line 1 (denoted by `m1`) or on line 2 (`m2`) only, or it has to go through the line 1 and then through line 2 (`m12`) or in reverse order (`m21`). As an example, the token (`1,m12`) refers to the workpiece having identifier '1' and this piece has to be processed first on line 1 then on line 2. The state of input and output of manufacturing lines is modelled with two places. The places `Tin_empty`, `Tin_full` refer to state of inputs and they are mutually exclusive. The colour of tokens referring to the state of these places consists of the identifier of the line only. If the input place of the line 1 is empty then there is a token having colour `m1` on the place `Tin_empty`, and there has not to be a token having colour `m1` on the place `Tin_full`. The same applies for the places (`Tout_empty` and `Tout_full`) describing the state of the output places of manufacturing lines.
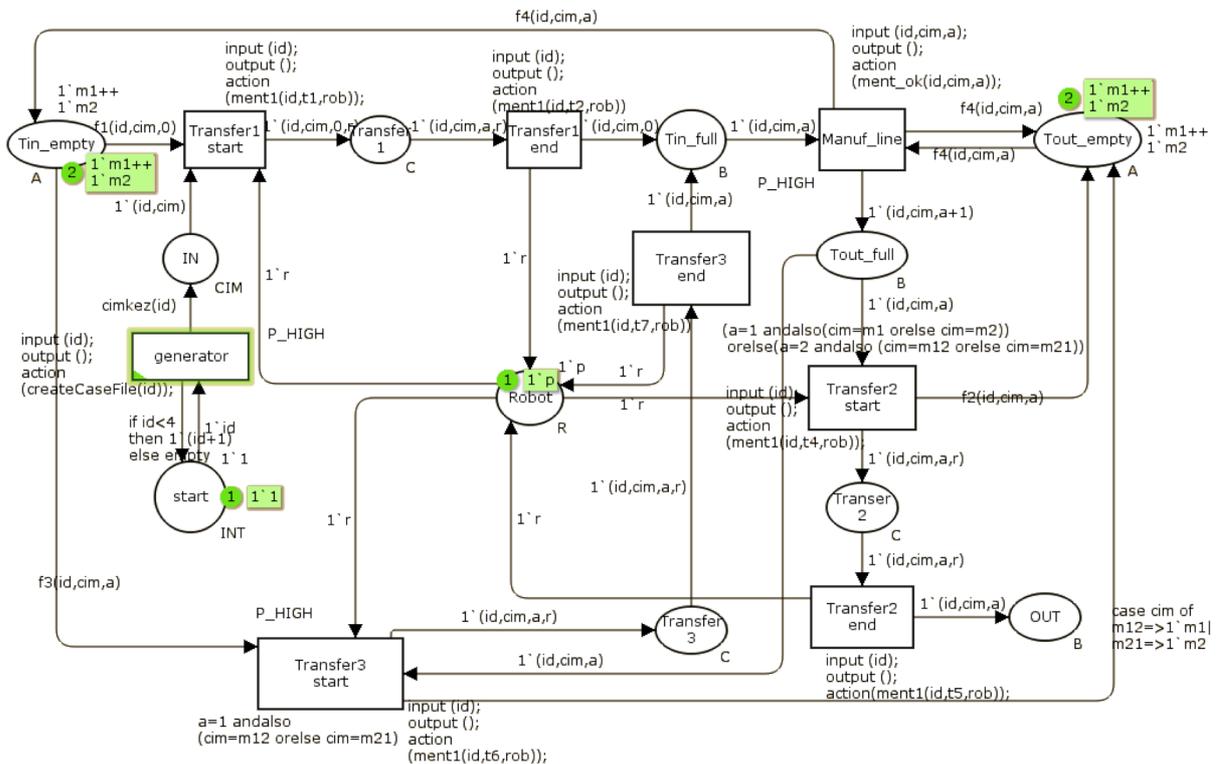
Figure 2: The CP-net model of the faultless operation of the manufacturing system

The transfer processes are disintegrated into three steps: to transitions referring to the (i) start and (ii) end of the transfer and to a place representing the (ii) transfer itself. Transitions `Transfer#_start` correspond to the start of transfer processes from `IN`, if `# = 1`, from `Tout_full` if `# = 2` or 3. The places `Transfer#` represent the transfer process 1, 2 or 3. The transitions `Transfer#_end` refer to completing of transfer process to the input place of a line (`# = 1`), to product container (`# = 2`) or to the input place to another manufacturing line (`# = 3`). The place `Manuf_line` refers to two manufacturing lines and the colour of token shows the line being processed.

Assume that only a single fault can occur in the system during manufacturing: the identification label of the piece can get damaged therefore it cannot be identified. Workpieces with damaged label get into a separate container represented by place `OUT_Fault`. The repairing of the label is not handled in this example. The modified part of the Petri net model can be seen in *Fig.3,* where the occurring of fault is taken into account. The occurrence of the fault is forced by a check function built into arc inscriptions in the Petri net model randomly. This check function returns with a fault in predefined probability. This probability of the can be set in the definition part of the net and different fault probability values can be assigned to the two manufacturing lines.

By comparing *Figs.2* and *3*, it can be stated that a new place `OUT_fault` appears as a consequence of fault modelling and integration and the arc expression functions of arcs starting from transition `Manuf_line` are extended with the fault checking.

For realistic investigation of a technological system the timed version of the CP-net model is used as a case study. A time point is assigned to the transitions. The transition `generator` does not belong to the technological system closely, so it fires under zero time, i.e. instantaneously. Different time units are assigned to the other transitions. These time values appear as transition inscriptions '`@+i`' (where `i` is an integer number defining the amount of time in seconds) on the net as it can be seen in the *Fig. 3*. During the simulation investigations different time values have been applied in order to check the possibility of a deadlock.

Preliminary analysis by simulation

As a first step, simulation is applied for the investigation of the developed Petri net model of the manufacturing system. These were carried out assuming both faultless system operation and when fault can occur during the manufacturing. Both non-timed and timed operational modes were considered.

The short description of the simulation is as follows. The simulation starts with the generation of tokens representing workpieces. The number of these tokens i.e. the number of workpieces to be processed can be modified in the arc expression function belonging to transition `generator`. The type of processes to be
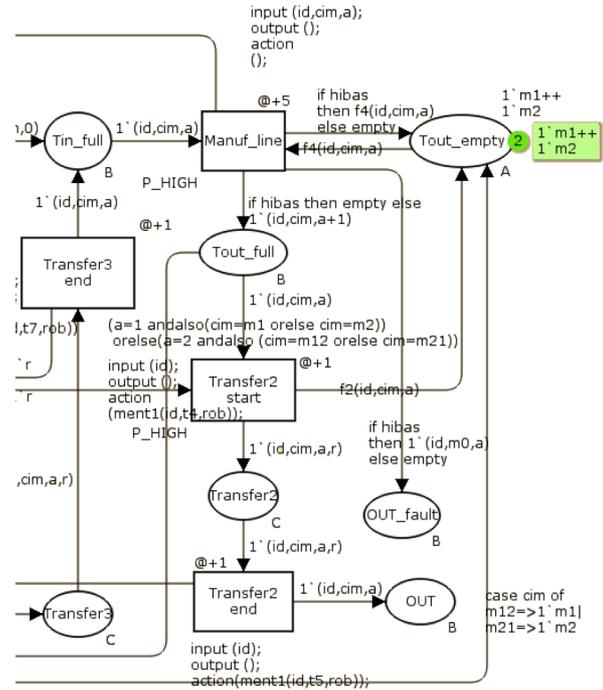
Figure 3: The Petri net model after integration of fault

performed i.e. the label referring to the manufacturing line(s) of the tokens can be set in the definition part of the CP-net. As a result, the colour of generated tokens refers to the identifier and to the process(es) to be carried out. Because of the highest priority of transition `generator` the transition `Transfer1_start` can only fire after the prescribed number of token appears at the place `IN`. As mentioned above, the transition `Transfer1_end` refers to the completion of the transfer of the workpiece to the input place of a manufacturing line. As a next step, the firing of transition `Manuf_line` corresponds to the completion of manufacturing process. The only fault in the system can occur during this process. If it happens then the token gets into the place `OUT_fault`, which represents the fault container. If the fault does not occur then the system checks whether the manufacturing process is over or the workpiece has to be also processed on the other line. In the first case the robot puts it to the place `OUT` representing the product container, while in other case it transfers the token representing the piece into the input place of another line. The modelling of the other two transfer processes (transfer of a workpiece from the output place of a manufacturing line either to the product container or to the input place of the other line) is similar to the transfer from place `IN` to place `Tin_full`. If the appropriate input place and the robot are free, another transfer process can start.

The primary goal of the simulation is to check the correct operation of the model. Another aim is the investigation of possible deadlock situations. In case of deadlock, the process stops without completing all the prescribed technological actions. It can happen in case of timed simulation of our investigated system and it refers to the wrong determination of the timing of actions.
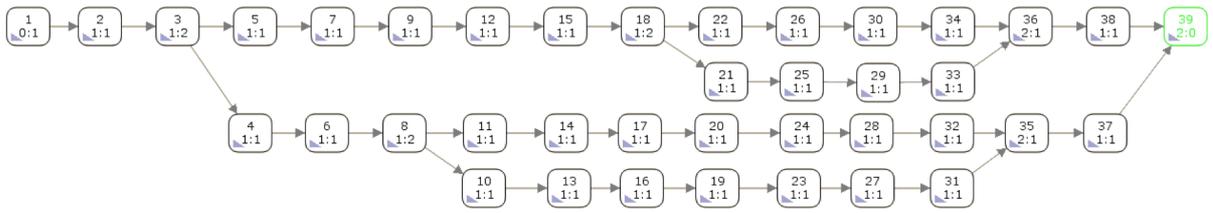
Figure 4: The occurrence graph - no fault, no timing

token distribution
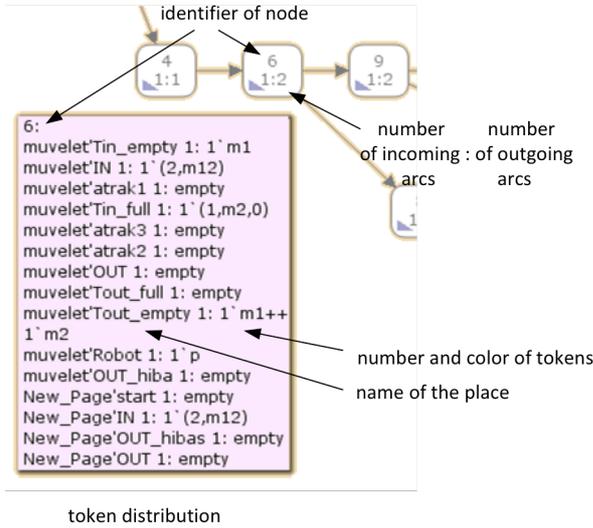
Figure 5: Key for the occurrence graph

The CPNTools generates automatically the occurrence graph, but the check functions used for the fault generation have to be removed from the arc expressions otherwise the occurrence graph is generated only for the normal mode or for the faulty mode.

For the illustration of generation and analysis of the occurrence graph let us assume that there are two workpieces to be processed, one of them has to go through manufacturing line 2, while the other has to go first through manufacturing line 1 then through line 2. Let the operation of the system be faultless and let the firing of all transition be instantaneous, i.e. the net is non-timed. The resulted occurrence graph can be seen in *Fig.4*. The explanation of numbers in the occurrence graph can be seen in *Fig.5*. The frames in the upper part of boxes are the identifiers of system states. The expression *x:y* in lower part shows the number of preceding and succeeding states. The token distribution belonging to a node can be obtained by selecting the triangle in left low corner. The opening window contains the name of places and the number and colour of tokens.

All the branches on the graph are explained by technological reasons, as there is no built in rule for the robot to start the transfer with any particular workpiece. The graph has only one terminal node (highlighted by green) and it refers to the normal termination of the process i.e. all the prescribed manufacturing processes terminated properly.

Now we can repeat the simulation with the same initial condition relating to workpieces but assuming that faults can occur during manufacturing process. The resulted occurrence graph can be seen in *Fig.6*.

Simulation cannot ensure the thorough investigation of the modelled system, but it complements well the further analysis. For the detailed investigations we applied the analysis of occurrence graph.

Analysis based on the occurrence graph

The thorough analysis of the behavioural properties of a CP-net can be performed using its occurrence graph. The concept of the occurrence graph was introduced above.

Figure 6: The occurrence graph - with fault, no timing
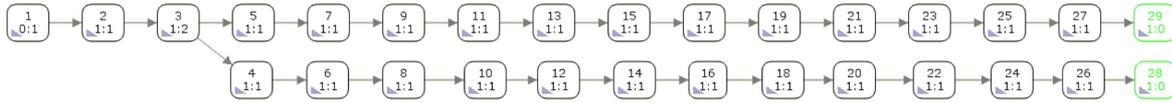
Figure 7: The occurrence graph - no fault, with timing
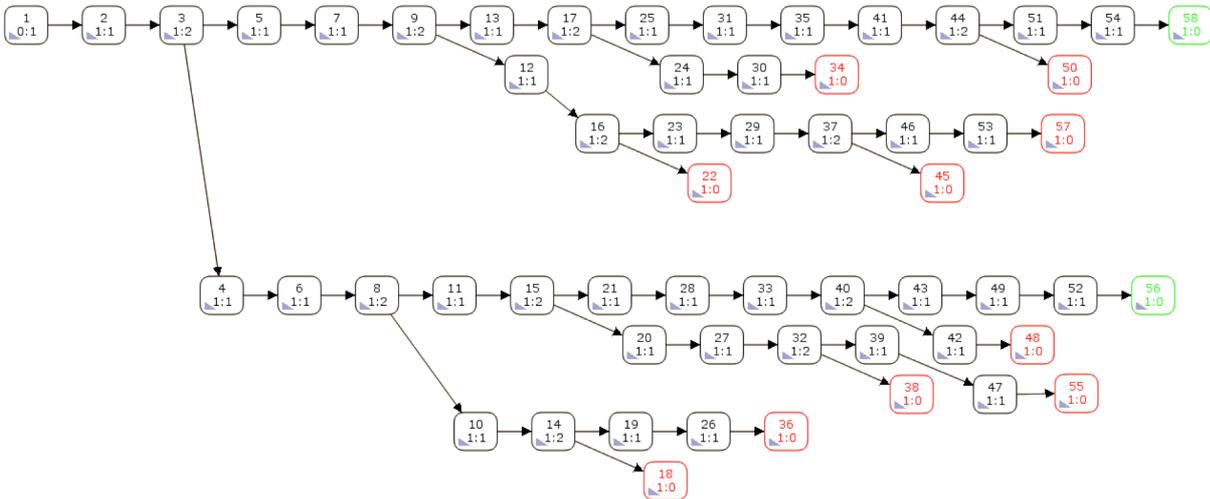
Figure 8: The occurrence graph - with fault and timing
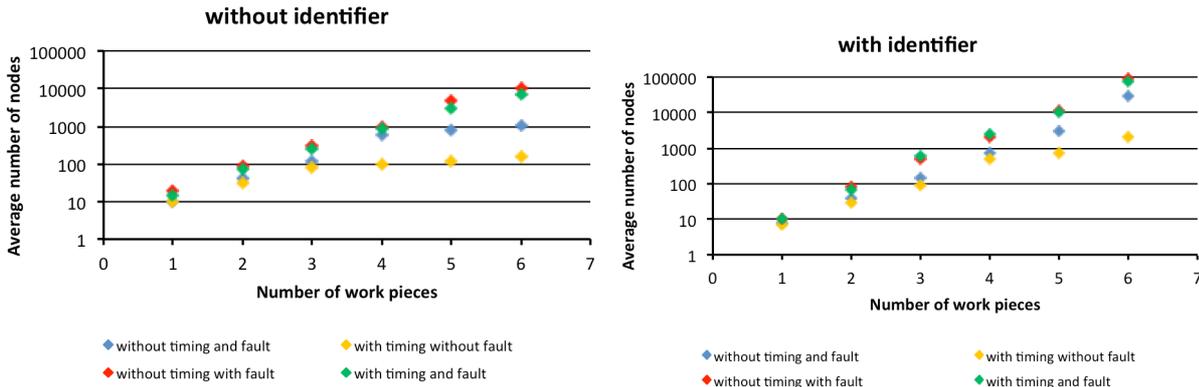
Figure 9: The node numbers in occurrence graph

It can be stated that the occurrence graph has become larger and the number of terminal nodes has increased due the effect of the possible faults. The reason of branches on the graph can be either technological (at the nodes 3, 9, 28, 51 and 54) or due to the fault. Only one of the terminal nodes refers to the normal termination of the process (highlighted by green, node 62), the others belong to the faulty cases (highlighted by red, nodes 22, 37, 43, 56 and 57). In case of faulty operation the identification tag of at least one workpiece get damaged during the manufacturing process.

Next we repeat the simulation again but adding the timing information to the net and assuming faultless operation. Let the time duration of transfer processes be equal to unit time, and the time of manufacturing processes is equal to 5 time units. The resulted occurrence graph can be seen in *Fig.7*.

The resulted occurrence graph is a tree and its two terminal nodes (nodes 28 and 29) differ only in time

stamp of the tokens (the time stamp can be seen only in CPNTools with a detailed label for each node). Comparing this graph with the occurrence graph in *Fig.4*, it can be stated that the number of parallel branches is less and they have different terminal states. Adding timing information to the model reduces the number of the possible different technological variants.

As the fourth case, let the simulation be performed with both timing and faults. The resulted occurrence graph (see *Fig.8*) is a tree again, the number of nodes is almost the same but the number of terminal nodes has doubled comparing to the occurrence graph in *Fig.6*.

It can be stated based on *Figs.4* and *6–8* that the occurrence graph is relatively simple in case of small number of workpieces and the analysis of nodes can be done manually. Thus, it is easily to find the terminal node or nodes referring to normal, faultless termination of the process, and those terminal nodes where the manufacturing of one piece or of both pieces ends with fault. However, the size of the occurrence graph grows

Table 1: Comparing the structure of occurrence graphs – no timing

No. of work-pieces	label (identifier, manufacturing information)	faultless operational mode			faulty operational mode		
		nodes	arcs	TN[a]	nodes	arcs	TN[a]
1	(1,m1)	7	6	1	8	7	2
1	(1,m12)	10	9	1	12	11	3
2	(1,m1)++(2,m2)	29	32	1	42	50	4
2	(1,m1)++(2,m12)	37	40	1	60	70	6
2	(1,m2)++(2,m12)	39	41	1	62	72	6
2	(1,m12)++(2,m21)	42	44	1	82	95	9
3	(1,m1)++(2,m2)++(3,m12)	145	173	1	306	409	12
3	(1,m2)++(2,m12)++(3,m21)	165	189	1	414	536	18
3	(1,m1)++(2,m1)++(3,m1)	107	123	1	210	273	8
4	(1,m1)++(2,m2)++(3,m12)++(4,m21)	631	772	1	2063	2900	36
4	(1,m1)++(2,m1)++(3,m1)++(4,m1)	340	404	1	949	1300	16
5	(1,m2)++(2,m12)++(3,m21)++(4,m1)++(5,m2)	2208	2797	1	9813	14420	72
5	(1,m21)++(2,m12)++(3,m21)++(4,m12)++(5,m21)	3259	3839	1	23709	32939	243
5	(1,m2)++(2,m1)++(3,m2)++(4,m1)++(5,m2)	1621	2165	1	5252	8051	32
6	(1,m2)++(2,m12)++(3,m21)++(4,m1)++(5,m2)++(6,m12)	8698	10980	1	62751	93323	216
6	(1,m21)++(2,m12)++(3,m21)++(4,m12)++(5,m21)++(6,m12)	12184	14562	1	146992	208602	729
6	(1,m2)++(2,m1)++(3,m2)++(4,m1)++(5,m2)++(6,m1)	5638	7686	1	21630	38886	64

[a] terminal nodes

exponentially if the number of pieces becomes larger, as it can be seen in the left part of *Fig.9* assuming faultless or faulty operational mode, and non-timed or timed net.

In case of large number of simpler workpieces when the identification is not necessary for each item, the identification tag can be omitted from the colour of the token. Assuming this situation all the simulation investigations (with and without fault, with and without timing) is repeated. The size of the resulted occurrence graphs is less than an order of magnitude simpler compared to the equivalent case with identifier as it can be seen in the right part of *Fig.9*. Since the workpieces have to be processed on the same manufacturing line, they have the same colour, thus they are indistinguishable. Since the robot selects among workpieces having the same colour, there will be no technological branches on the occurrence graph. This results in a much simpler occurrence graph especially when there are large number of workpieces and the processes to be performed is one or two kind of sorts.

As an example, let us investigate the structure of occurrence graphs in case when there is no timing in the system, the token colour contains the identification tag and both faultless and faulty operational mode is assumed. According to data in *Table 1,* the size of the occurrence graph depends on the number of workpieces, the number of manufacturing procedures to be processed, and the presence of fault. If fault can occur during the manufacturing process both the number of nodes and the arcs increases dramatically. There are several terminal nodes, too, but only one refers to normal termination of manufacturing processes the others belong to different faulty situations. The labelling of workpieces has also a significant effect on the complexity of the occurrence graph. If there are one or more workpieces, which should be processed on both manufacturing lines then the number of nodes and arcs is doubled as it can be seen in the corresponding rows in *Table 1*.

The results of an investigation with timing information can be seen in *Table 2*. If the number of workpieces is less than four or the workpieces have to be processed on only one manufacturing line, then the size of the occurrence graph depends on the number of workpieces and the presence of fault. On the other hand if there are at least four workpieces and at least two of them have to be processed on both manufacturing lines, but in reverse order then a deadlock situation can occur. A further condition of a deadlock that the manufacturing time should be longer than the transfer time but it is true in general.

In case of a deadlock, the process stops because the precondition of transfer processes cannot be fulfilled. There are workpieces on input and output places of both manufacturing lines and therefore no further steps are enabled. If all the workpieces have to be processed on both manufacturing lines and there is no fault then all of terminal nodes refer to a deadlock as it can be seen in the rows marked by a footnote in the *Table 2*.

The identification tag can be omitted from the colour of token in certain cases. If the number of workpieces is small and they have to be processed in different ways then there is a significant change in the occurrence graph. On the other hand in case of large number of workpieces, the structure of occurrence graph becomes much simpler if the identification tag is removed from the colour.

Analysis of the occurrence graph using the OGAnalyzer

As mentioned above, CPNTools cannot use the information about the probability of faults at the generation of occurrence graph. However, assigning this value to the appropriate edges, the probability of each node of the occurrence graph, i.e. of each system state can be determined. For this purpose, software called

Table 2: Comparing the structure of occurrence graphs – with timing

No. of work-pieces	label (identifier, manufacturing information)	faultless operational mode			faulty operational mode		
		nodes	arcs	TN/D[a]	nodes	arcs	TN/D[a]
1	(1,m1)	7	6	1/0	8	7	2/0
1	(1,m12)	10	9	1/0	12	11	3/0
2	(1,m1)++(2,m2)	23	22	2/0	41	40	8/0
2	(1,m1)++(2,m12)	29	28	2/0	59	58	12/0
2	(1,m2)++(2,m12)	29	28	2/0	58	57	12/0
2	(1,m12)++(2,m21)	35	34	2/0	87	86	18/0
3	(1,m1)++(2,m2)++(3,m12)	73	74	4/0	264	263	48/0
3	(1,m2)++(2,m12)++(3,m21)	86	85	4/0	392	391	72/0
3	(1,m1)++(2,m1)++(3,m1)	85	84	6/0	262	261	48/0
4	(1,m1)++(2,m2)++(3,m12)++(4,m21)	173	172	8/2	1473	1472	258/2
4	(1,m1)++(2,m1)++(3,m1)++(4,m1)	365	364	24/0	2121	2120	384/0
5	(1,m2)++(2,m12)++(3,m21)++(4,m1)++(5,m2)	574	573	24/4	9320	9375	1582/8
5[b]	(1,m21)++(2,m12)++(3,m21)++(4,m12)++(5,m21)	154	153	24/24	12076	12075	2196/60
5	(1,m2)++(2,m1)++(3,m2)++(4,m1)++(5,m2)	514	513	24/0	5152	5247	780/0
6	(1,m2)++(2,m12)++(3,m21)++(4,m1)++(5,m2)++(6,m12)	1725	1732	80/40	71817	72336	12300/116
6[b]	(1,m21)++(2,m12)++(3,m21)++(4,m12)++(5,m21)++(6,m12)	325	324	72/72	45541	45540	8280/432
6	(1,m2)++(2,m1)++(3,m2)++(4,m1)++(5,m2)++(6,m1)	2341	2412	72/0	39349	41076	5760/0

[a] terminal nodes/deadlocks, [b] labels where all terminal nodes refer to deadlocks

OGAnalyzer has been developed. We assumed that the occurrence graph belonging to a given initial state of a CP-net model is finite and acyclic. The occurrence graphs of Petri nets modelling manufacturing systems fulfil this assumption in general.

Let the probability of faults be known from technological consideration and let the first step of the analysis be the assignment of the arc weights to the edges of occurrence graph as follows.

1. If a node on the occurrence graph has only one outgoing arc, then the next state follows unambiguously, thus the arc weight is equal to 1.
2. If there is more than one outgoing arcs from a given node then it means that different states can follow from it. These states come into existence with different probabilities depending on whether this branch has technological or fault related reason.
   a. If there is no fault in the system then every branch has a technological reason, because there is no built-in priority rule for the robot to the selection among the workpieces. It means that every selection that is every arc has the same probability so all of the arcs starting form this node have to get the same arc weight, which is equal to the reciprocal value of the number of outgoing arcs.
   b. The introduction of the fault into the model causes the appearance of another type of branching in the occurrence graph. Let us assume that only one type of fault can occur in a given system state. It results in 2 new system states: one for the normal operation and one for the faulty mode. Let the probability of fault be equal to P_f. Then the weight of arc leading to faulty mode is equal to the probability of the fault while the arc leading to the normal operational mode gets the value $1-P_f$.

Assigning these arc weights to the edges of the occurrence graph the probability of a given state on the graph can be determined in the following way if the faults occurring one after the other in the system are independent:

1. If the graph is a tree or only one route leads to the given node then the probability of the state representing by this node is equal to the product of arc weights along the route leading from the node representing the initial state to the given node.
2. If there are more than one route leading to the given node from the initial node then the probability value of each route has to be determined with the production of arc weights along the route as in the first step, then to sum these resulted values.
3. If the faults are not independent from each other, then the probability of nodes can be calculated in a similar way but using conditional probability values.

As described above, this determination method and the operation of OGAnalyzer has been illustrated using the example of two workpieces to be processed and one of them has to go through manufacturing line 2, while the other has to go first through line 1 then trough line 2. Faults can occur during the manufacturing process and let the fault probability be equal to 0.3 in case of line 1 and 0.1 in case of line 2. As before, we consider the timed case, when the transfer transitions have the same transition time of 1 time unit, while the manufacturing time is equal to 5 for both lines. After the simulation CPNTools generates the occurrence graph, the structure of, which is the input of OGAnalyzer.

As a first step OGAnalyzer reads the structure of the generated occurrence graph from the data file generated by CPNTools and visualizes it in its own window as it is shown in *Fig.10*. The user can get the token distribution belonging to nodes as it is shown at node 1 in *Fig.10*. The next step is the identification of branches. The software can distinguish between the two different types of branches on the occurrence graph (that are technological and fault caused branches). For this the user has to define the fault colour for the appropriate branches in a separate window.
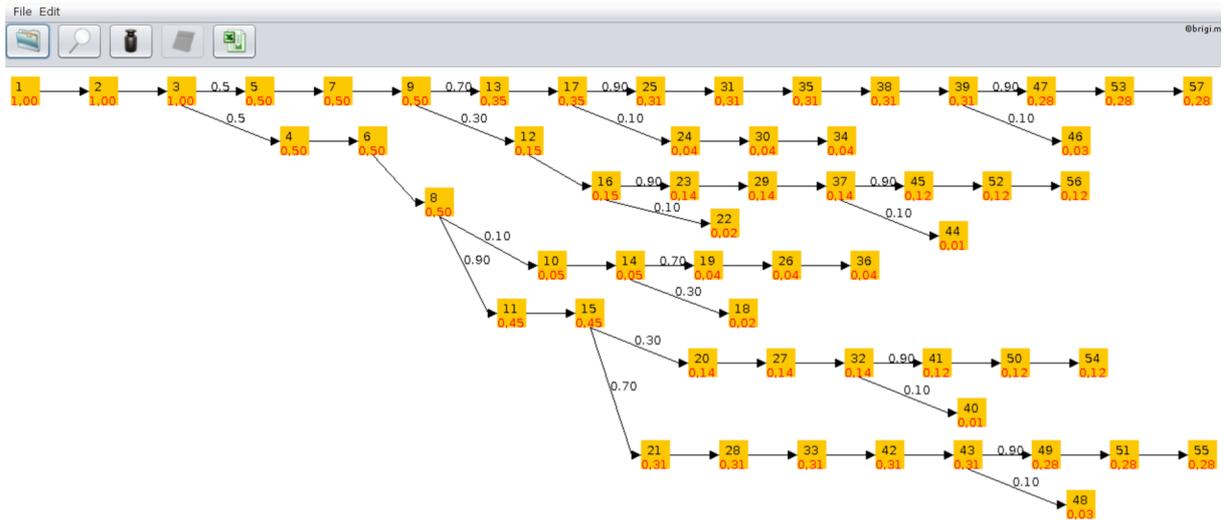
Figure 10: The occurrence graph of the example in the window of *OGAnalyzer*

It is assumed that the technological branches have the same probability, so OGAnalyzer assigns the reciprocal of the number of branches to these edges. The fault caused branches are collected into a table and the user has to define the probability of faulty and normal modes as it can be seen in *Fig.11*. The occurrence graph with arc weights depicted in OGAnalyzer can be seen in *Fig.12*. The black numbers attached to the arcs are the arc weights. Values equal to 1 were not depicted.

Figure 11: Defining the probabilities values of faults

Using these probability values the OGAnalyzer calculates automatically the probability of a node by multiplying arc weights along the path leading from the node representing the initial state to that node. If two or more paths lead to the node then the probability values belonging to these paths are summed. The calculated probability values can be also seen in *Fig.12* as red numbers assigned to nodes.

In this manner the probability of all system states in the investigated system can be calculated. For example, the probability of faultless completing of both workpieces is 0.28 + 0.28 = 0.56, which is equal to the sum of the probabilities belonging to the nodes 55 and 57. The probability of that case when the first workpiece is manufactured without fault, but the label of the other piece gets damaged during the second manufacturing process is 0.03 + 0.03 = 0.06 (sum of the probabilities belonging to the nodes 46 and 48).

In case of timed nets the occurrence graph is often a tree and the same token distribution belongs to different nodes because of the different time stamp. For the determination of probability of a given system state, the probability values belonging to different nodes have to be summed up.
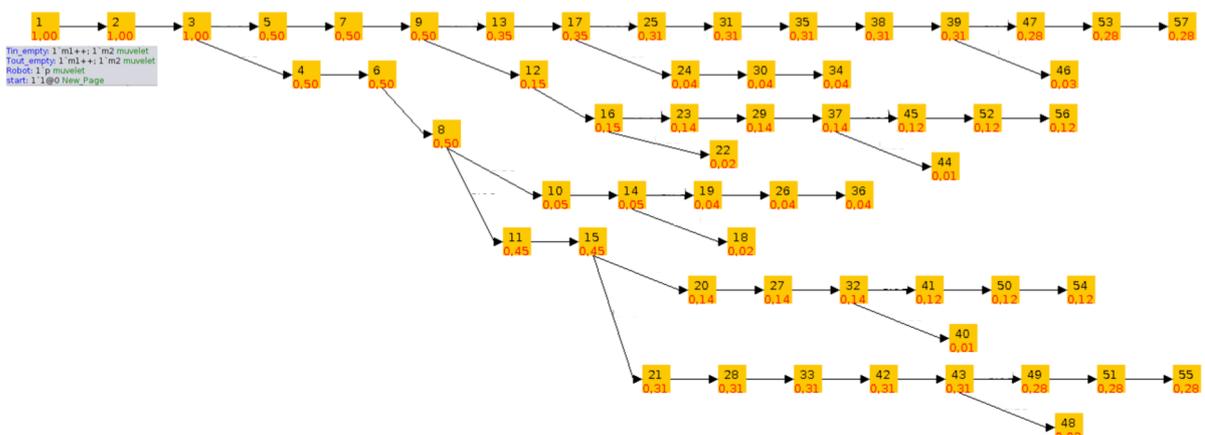
Figure 12: The occurrence graph with arc weights and probability values

Conclusion

A novel occurrence graph investigation procedure for discrete event systems described by Petri nets was proposed in this paper for model-based diagnostic purposes that utilize the knowledge of the occurrence probability of faults. The model of the investigated system was defined in timed coloured Petri net form. The colours of tokens representing the workpieces were used to distinguish them and to assign a label of the processes to be carried out. The arc inscriptions and built-in probability functions were used for the fault modelling and integration.

The operation of the system was investigated via simulation both in timed and non-timed cases and in faultless and possible fault operational modes with given number pieces and prescribed manufacturing lines.

For the behavioural analysis of the model the occurrence graph method was used. A special software module, called OGAnalyzer has been developed for the handling of the probabilities on the occurrence graph and for calculating the occurrence probability of system states.

Acknowledgements

REFERENCES

[1] CASSANDRAS C.G., LAFORTUNE S.: Introduction to Discrete Event Systems, Kluwer Academic Publishers, *1999*

[2] GERZSON M., MÁRCZI B., LEITOLD A.: Diagnosis of Technological Systems based on their Coloured Petri Net Model, ARGESIM Report no. S38 (Eds. TROCH I., BREITENECKER F.) *2012*, p. 358/1–6

[3] JENSEN K.: Coloured Petri Nets: Basic Concepts, Analysis Methods and Practical Use, Springer-Verlag, *1997*

[4] CPN GROUP, University of Aarhus, Denmark: CPNTools 2.2.0 http://wiki.daimi.au.dk/cpntools/ (last accessed: May 25, 2014)

[5] FANTI M.P., SEATZU C.: Fault diagnosis and identification of discrete event systems using Petri nets, Proc. 9th International Workshop on Discrete Event Systems, WODES, *2008*, 432–435

[6] JENSEN K., KRISTENSEN L.M., WELLS L.: Coloured Petri Nets and CPN Tools for Modelling and Validation of Concurrent Systems, Int. J. of Software Tools for Technology Transfer, *2007*, 9(3–4), 213–254

[7] MURATA T.: Petri Nets: Properties, Analysis and Applications, Proceedings of the IEEE, *1989*, 77(4), 541–580

IMMEDIATE EVENT-AWARE MODEL AND ALGORITHM OF A GENERAL SCHEDULER

Tibor Dulai[1][✉], Ágnes Werner-Stark[1], and Katalin M. Hangos[1,2]

[1] Department of Electrical Engineering and Information Systems, University of Pannonia, Egyetem str. 10, Veszprém, H-8200, HUNGARY
[2] Process Control Research Group, Computer and Automation Research Institute, Kende u. 13-17, Budapest, H-1111 HUNGARY
[✉]E-mail: dulai.tibor@virt.uni-pannon.hu

A stochastic scheduling problem is investigated in this work that considers workpieces to be manufactured according to individual recipes containing manufacturing steps performed by workstations as resources. Unexpected stochastic breakdown of a workstation or the faulty termination of a recipe, when a manufacturing failure renders the workpiece out of specifications, forms the set of immediate events. A model and an algorithm are proposed as the basis of a scheduler, which takes into account the possible immediate events, estimates their probability and suggests resource allocations which provide the best overall work-flow even when an immediate event happens. This model includes the possibility of handling alternative resources that can substitute each other in case of an immediate event, like sudden technical failure of a resource. Immediate events are not exactly predictable; however, based on previous experiences, their probabilities can be estimated. Our model uses the properties of the resources (including how they can substitute other types of resources) and the required sequence of them during the workflow (i.e. the recipes). The proposed scheduling algorithm constructs a solution workflow that reacts in the best way (in average) even for an unexpected event. The proposed model and scheduling algorithm is illustrated on two industrial case studies.

Keywords: scheduling, resource allocation, alternative resources, immediate event-awareness

Introduction

Scheduling is an important and widely used topic of operations research. Besides of its theoretical importance, industries can also benefit from optimal schedules and resource allocation. Several different algorithms were established for organizing process elements on the time scale [1, 2], related to for example computer networks [3], business processes [4, 5] or industrial processes [6]. The results are usually represented on a Gantt chart.

There are also differences between these algorithms related to their application times. Some of them are applied offline before the scheduled processes starts, others are applied real-time. Real time scheduling techniques, and a special adaptive real-time scheduler is introduced in Ref. [10]. For improving their results, there are cases when the methods use historical data during the creation of the schedule [11].

The common scheduling methods usually handle the resources individually and do not take into account the relationships, e.g. the similar functionality between them. Only few publications investigate the cooperation possibilities that are enabled by the similar functionality of resources in scheduling [7].

In this paper a scheduling method is proposed, where resources may substitute each other and immediate failures of resources may happen. After we introduce the problem and its main building blocks, we present our model and the algorithm developed. Its operation and properties are demonstrated on a simple and a more complex problem as case studies.

Problem specification

A general scheduler intends to determine the placement of activities of resources on the time scale. In an advanced case one might consider additional features given in the problem's model – like substitution possibility of the resources – which help to redefine the classic scheduling problem to be usable in different real-life applications. This substitution may help in achieving a certain fault tolerance property in such a way that a technical failure causes the least possible negative effect. The basic aspects and sub-tasks related to this extended scheduling problem are collected in this section.

Cooperation

In some cases more than one resource is able to carry out a particular activity (usually with different productivity). It means that when the appropriate resource is busy, there can be another resource that is able to take over a particular task if required. We shall term this case a cooperative situation when substitution is possible, In other words, the resources can cooperate with each other. Taking the cooperation aspect into account during scheduling the performance may increase [8]. We introduced the so called substitution vector as an element of our model to handle substitution related sub-problems.

Any scheduler [1, 2] can be applied as a basis of the cooperative extension, which inputs the sequence of resources as input processes and has to arrange them on the time scale taking into account the availability and temporal constraints. All the improvements are done on this basic schedule.

Improving fault tolerance

When a schedule is ready and the workflow starts, an immediate event (e.g. a technical failure of a resource) in a process may influence other processes, too, in a negative way [9]. If the effects of that event are calculated in the schedule, then the solution with the best answer to immediate events can be selected from the set of solutions with the same performance, and the faults' negative impact on the schedule decreases. Our model deals also with this aspect.

The model

For the above mentioned problem set, we created an universal model and an algorithm to carry out the scheduling. The work intends to be the basis of several different tasks with the goal that the operation time should be minimal. In this section we introduce our model and the scheduler that works on this model.

Main parameters, notations and functions

The model and its parameters are designed in such a way that it can be applied for different kinds of scheduling problems (e.g. scheduling of industrial production processes, test processes, scheduling and managing the resources of electrical networks, etc.). For this reason we collected the necessary parameters that make possible the development of a general framework, which uses different kinds of resources in different processes. It is able to handle different needs. In this sub-section we introduce the main parameters and functions we use in our work. These are as follows:

$$\textbf{Process} = \{proc_1, \ldots, proc_j\}, \tag{1}$$

stands for the set of processes;

$$\textbf{P} = \{p_1, \ldots, p_l\}, \tag{2}$$

stands for the set of product types;

$$\textbf{R} = \{r_1, \ldots, r_m\}, \tag{3}$$

stands for the set of resources;

$$\textbf{O}(r_i) = \{o_{i1}, \ldots, o_{ik}\}, \tag{4}$$

stands for the set of the operation modes of resource i;

$$\textbf{A} = \{a_1, \ldots, a_n, pause(t)\}, \tag{5}$$

stands for the set of basic activities, where *pause(t)* is an empty activity with length of t hour;

$$ra : \textbf{R} \times \textbf{A} \rightarrow \{0, 1\}, \tag{6}$$

defines a function for determining whether a resource is able to perform an activity;

$$rap: \textbf{R} \times \textbf{A} \times \textbf{P} \rightarrow \{0, 1\}, \tag{7}$$

defines a function for determining whether a special activity of a resource can be applied for a product type;

$$t: \textbf{R} \times \textbf{A} \times \textbf{P} \times \textbf{O} \rightarrow \textbf{N}, \tag{8}$$

provides the suggested operation time of a resource in a given operation mode performing a given activity on a given product type;

$$q: \textbf{R} \times \textbf{A} \times \textbf{P} \times \textbf{O} \rightarrow [0,100], \tag{9}$$

provides capacity information: how many percent of a resource capacity is occupied by one piece of a given product type in a given operation mode of a resource while performing a given activity;

$$e: \textbf{R} \times \textbf{A} \times \textbf{P} \times \textbf{O} \rightarrow [0,100], \tag{10}$$

provides the probability of resource failure during performing a given activity on a given product type in a given operation mode;

$$s: \textbf{R} \times \textbf{A} \times \textbf{P} \times \textbf{O} \rightarrow \{0, 1\}, \tag{11}$$

results in a binary decision: whether a resource activity in a given operation mode on a given product type can be suspended without restarting it from its beginning;

$$rreq: \textbf{R} \times \textbf{A} \times \textbf{P} \times \textbf{O} \times \textbf{DateTime} \rightarrow \textbf{P(R, N)} \tag{12}$$

(power set on pairs of a resource and a natural number), provides the additional resource need of a resource's given activity on a given product type in a given operation mode in a given hour

$$subst(r_i, a_j, p_k, o_{il}) = [\ h_{i1}\ \ldots\ h_{im}\ ], \tag{13}$$

is the substitution vector, where m is the number of resources, h_{in} is a natural numbers for all $1 \leq n \leq m$, r_i is the ith resource, a_j is a basic activity, p_k is a product type, o_{il} is an operation mode of resource i, h_{in} denotes how many percent of the productivity of resource i is needed in a given activity in a given operation mode on a given product type to substitute totally resource n supposing unchained operation time. The substitution vector is calculated based on the function ra and other functions (e.g. function t in case of time-based optimization).

$$map: \textbf{R} \text{ x } \textbf{DateTime} \rightarrow \textbf{N}, \qquad (14)$$

provides the information on accessibility of resources: how many resource of a given type is accessible in a given hour;

$$\textbf{F}: \textbf{R} \rightarrow \textbf{N}, \qquad (15)$$

provides the expected number of hours how long a given resource is unavailable in case of its failure;

$$t_{start}: \textbf{Process} \rightarrow \textbf{DateTime}, \qquad (16)$$

shows the start time of a process;

$$t_{maxend}: \textbf{Process} \rightarrow \textbf{DateTime}, \qquad (17)$$

shows the maximum finish time of a process;

$$prev: \textbf{Process} \times \textbf{A} \times \textbf{N}^+ \rightarrow \{\textbf{A}, \emptyset\}, \qquad (18)$$

provides the prior basic activity of the nth occurrence of a given basic activity in a given process;

$$next: \textbf{Process} \times \textbf{A} \times \textbf{N}^+ \rightarrow \{\textbf{A}, \emptyset\}, \qquad (19)$$

provides the following basic activity of the nth occurrence of a given basic activity in a given process;

$$maxdelay: \textbf{Process} \times \textbf{A} \times \textbf{N}^+ \rightarrow \textbf{N}, \qquad (20)$$

provides the maximal duration of time out, which is tolerated by the nth occurrence of a given basic activity in a given process;

$$dur: \textbf{Process} \times \textbf{A} \times \textbf{N}^+ \rightarrow \textbf{Q}, \qquad (21)$$

provides the time scale, which number the default operation time of the of the nth occurrence of a given basic activity in a given process has to be multiplied with, for getting the real operation time of the activity.

The proposed algorithm

Our algorithm was created to take into account the cooperation possibilities of the resources and to have failure-aware behaviour during scheduling. The
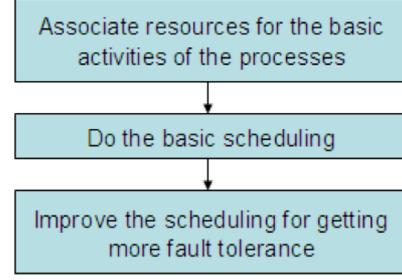
Figure 1: The basic parts of our algorithm

algorithm can be separated into three main blocks, as it is shown in *Fig.1*.

In the first phase, we associate resources for each basic activity of the processes. At the start, processes are described only as a sequence of basic activities. We have to turn it into the "language" of resources. In this task, we use the function ra for determining, which resources are able to carry out the desired basic activity. Furthermore, we deal with the minimization of the operation time; we select the resource, which has the operation mode with the minimal operation time, i.e. r_i with

$$\min(t(r_i, A, P, o_{il})), \forall o_{il} \in O(r_i), ra(r_i, A) = 1 \quad (22)$$

is selected.

The first step of our algorithm results in a sequence of resources for each process. The second phase of the algorithm is performed by the main scheduler task, taking into account the alternative resources.

The first task in this phase is to select the "basic" process, which has the least robustness. We do it by selecting $proc_i$ with

$$\min(t_{maxend}(proc_i) - t_{start}(proc_i) - TDUR), \qquad (23)$$

where *TDUR* is the sum of the durations of all of the basic activities of $proc_i$.

After the determination of the basic process, the algorithm enters a loop, in which it selects a process from the set of the remaining processes and attempts to place all the basic activities of the selected process element-by-element on the time scale. During this operation the algorithm handles the substitutability of the resources. The insertion of a basic activity starts at the initial activity of the process and each activity is inserted into the earliest possible time point. This assumption is a fundamental point in the method. The insertion of an activity at its earliest time may have two possible outcomes:
- successful insertion
- unsuccessful insertion, which means that there is not enough amount of the resource to serve the activity which starts at its earliest time point. We call this case a collision.

In case of a collision, the algorithm attempts to find an alternative resource using the substitution vector,

Figure 2: Activity insertion onto the time scale

which has enough leftover capacity to carry out the activity at the desired earliest time point. If this is not possible, because there is not enough resource at the desired time, the algorithm tries to shift one of the collided activities to solve the problem. The detailed mechanism of the activity-insertion task is illustrated in *Fig.2*. When the activity-shift resulted in no success, the algorithm attempts to repeat the same process with other alternative resources. If there is no success, the whole process will be left from the final solution; otherwise the initial scheduling is ready.

In the final, third phase of our algorithm the schedule is tried to be modified in such a way that it should have a higher fault tolerance than the original one. This phase assumes that fault happens at the processing of special activities of the processes and calculates its effect into the schedule. We do not want to hurt the time constraints of the original problem that's why one of the main parameters of this phase are the maximum finish time values (t_{maxend}) of the processes. We process the processes starting from their end in the following way:

Let **WP** be the process list we work on. At the start **WP** = **Process**. For each element of **WP**, we create a pointer, which points at the penultimate activity of the process. The activities pointed by the pointers (let their set be **WA**) are candidates for assuming them to have fault. Starting from this initial state, this phase of the algorithm works as follows:

1. Select the a_j activity from **WA** with the highest error probability:

$$\max(e(r_i, a_j, p_k, o_{il})), a_j \in WA \qquad (24)$$

2. If the start of the basic activity which follows a_i can be shifted by a_i's duration plus by $F(R(a_i))$, where $R(a_i)$ is the resource which carries out a_i without causing collision and without exceeding $t_{maxend}(proc_k)$, where $proc_k$ is the process which belongs a_i to, then
(i) the shift will be done, and
(ii) the basic activity, which precedes a_i in the process will be added to **WA** and a_i will be taken out from there. It means that the pointer of the process will be set one step backward.

Figure 3: Activity sequence of the example processes

Else, a_i is taken out from **WA** and its process is taken out from **WP**, too. It means that the activities of the process are not modified any more.

This phase of the algorithm modifies the placement of the basic activities of the processes on the time scale in such a way that there is no need to change the scheduled activities in case of failures with high probability. It means that these failures will not cause delay and exceed of the maximum finish time.

Case studies

In this section, the previously presented algorithm is applied to two hypothetical manufacturing examples: we show how it takes into account the substitutability of the resources and how to make the schedule ready for possible technical failures.

A simple problem and its solution

Let there be four processes with the same start and maximum finish time ($t_{start} = 0$ and $t_{maxend} = 65$ are identical in each cases). The sequences of the process activities can be seen in *Fig.3*.

In this example, we work only with one product type (P), and we intend to minimize the maximum operation time. Let's suppose that in this example there are two pieces of resource r_1, while only 1 piece of the others, and

$$ra(r_i, a_j) = 1 \text{ only if } i=j, \qquad (25)$$

excepting two cases:

$$ra(r_2, a_3) = 1 \qquad (26)$$
and $$ra(r_3, a_2) = 1, \qquad (27)$$

as well. Moreover, the substitution vector of r_2 is:

$$subst(r_2, a_3, \mathbf{P}, \mathbf{O}) = [\ 0 \quad 100 \quad 125 \quad 0\ ], \qquad (28a)$$

while the substitution vector of r_3 is:

$$subst(r_3, a_2, \mathbf{P}, \mathbf{O}) = [\ 0 \quad 200 \quad 100 \quad 0\ ], \qquad (28b)$$

Figure 4: Sequence of the initial resource allocation

which means that r_2 can substitute r_3, and r_3 can substitute r_2. However, the substitution results in 125 percent or 200 percent more time, respectively.

Another assumption is when only r_1 has two operation modes: in one of its operation mode its operation time at a_1 activity on product **P** can be reduced with 10%. As our goal is to minimize the necessary time, we will use this operation mode of r_1 in each case. Moreover, assume that all the resources are unavailable for 2 time units in case of failure.

Based on this information, the first step of our algorithm creates the sequence of the resources for each process. This sequence completed with the durations of each activity after user-based modification of the default operation times is illustrated in *Fig. 4*. The second phase of the algorithm determines the basic process; in our case it is $proc_3$ with its 63 unit length. Its representation on the time scale is shown in *Fig.5*.

After this selection a loop is started, choosing the less robust process from the remaining set. The next

process is $proc_1$. Its first activity can be placed onto the time scale without collision. The placement of the second activity collides with the basic process's first activity. The first reaction of the algorithm is to search for alternative resource. Resource r_2 can substitute r_3 without collision; however in this case the operation time will increase from 4 to 5 units. The third activity of $proc_1$ can be inserted, since there are two pieces of resource r_1. The fourth activity collides again; however substitution solves this problem, and the problem-free placement of the final activity of the process can be seen in *Fig.6*.

The third process to handle is $proc_4$. Its first activity can be placed easily; however, the second activity has collision and there is no possibility to substitute it without collision either. This case requires the shift of the activity as illustrated by *Fig.2*. The activity that shifts the later start is reason for a pause was defined in $proc_4$. The third and fourth activities have to be also shifted. The difference between the two cases is the resource that is able to substitute gets free earlier. This is why the fourth activity will be carried out by r_2 instead of r_1, started after a short pause. After the placement of the process's final activity and all the activities of $proc_2$, this provided the schedule with some activities of the previously placed processes shifted necessarily twice, as illustrated in *Fig.7*.

The final phase of the algorithm is to make the schedule to be fault-aware. As $t_{maxend} = 65$ and the length of $proc_3$ is 64, we will not modify this process. Similarly, we will not modify $proc_1$ and $proc_4$ either. Only $proc_2$ lets the algorithm to prepare it to be fault-aware, and the modification can be applied to all of its
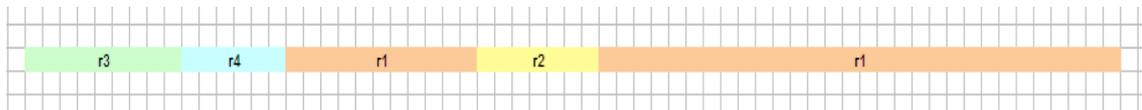
Figure 5: Placement of the basic process onto the time scale
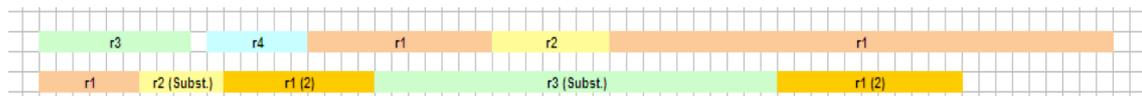
Figure 6: The schedule after handling *proc*1
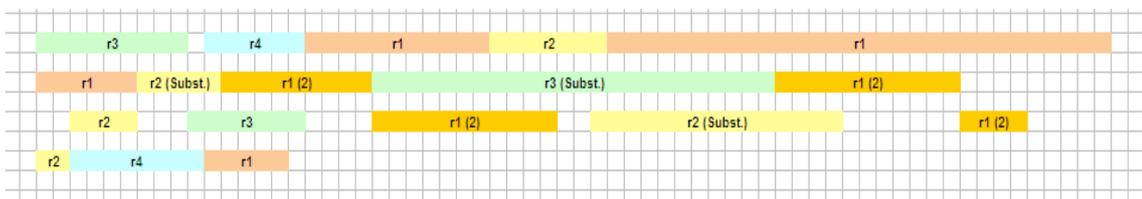
Figure 7: The schedule after the second phase of the algorithm
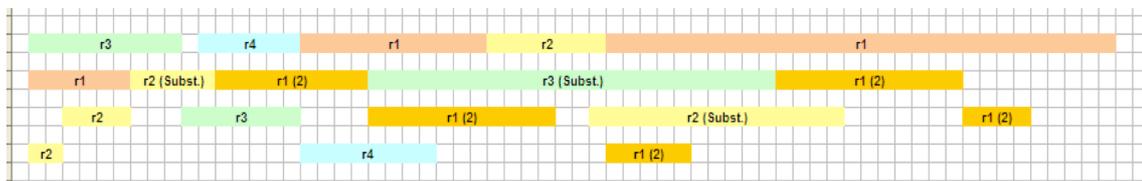
Figure 8: The final schedule

Table 1: Some properties of the processes of the second case study

process	start time, t_{start}	max. finish time, t_{maxend}
p_1	0	30
p_2	15	30
p_3	7	30
p_4	3	25
p_5	0	23
p_6	6	29

Table 2: The effect of failure on the second example's resources

Resource	$r1$	$r2$	$r3$	$r4$	$r5$	$r6$
Time of unavailability (F) [time unit]	2	1	3	2	2	1
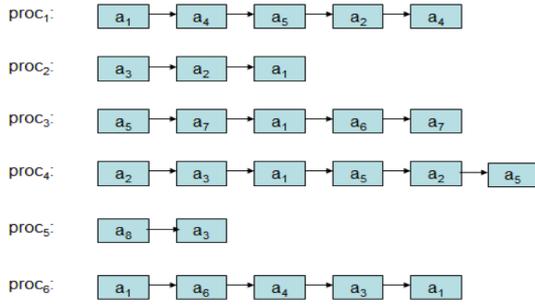
Figure 9: The activities of the second example's processes
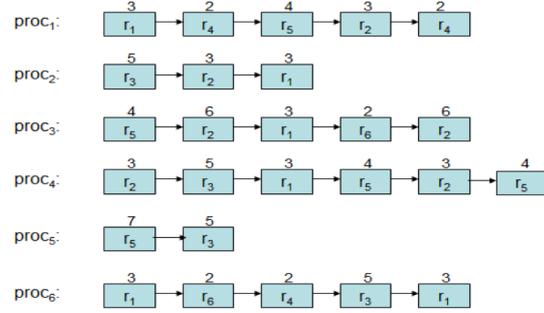
Figure 10: Resource-sequence of the example's processes

activities. The result can be seen in *Fig.8*.

A more complex problem and its solution

In this example we deal with six processes. In contrast with the previous example, they differ in their start and maximum finish time as shown in *Table 1*, where the values are represented in time units. The activity-sequences of the processes are illustrated in *Fig.9*.

In the example form *Fig.9*, eight different activities are applied. The next question is which resources are able to carry out these activities. In this example, there are six kinds of resources, six resource types (r_1 - r_6). Suppose that two pieces of resource type 2, 3 and 5 exist, while all the other resource types have only one representative. The substitution vectors of the resources are as follows:

$$subst(r_1, a_2, \mathbf{P}, \mathbf{O}) = [\ 100\quad 150\quad 0\quad 0\quad 0\quad 0\ ], \quad (29a)$$

$$subst(r_2, a_1, \mathbf{P}, \mathbf{O}) = \begin{bmatrix} 150 & 100 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad (29b)$$

$$subst(r_3, a_5, \mathbf{P}, \mathbf{O}) = \begin{bmatrix} 0 & 0 & 100 & 0 & 300 & 0 \end{bmatrix}, \quad (29c)$$

$$subst(r_5, a_3, \mathbf{P}, \mathbf{O}) = \begin{bmatrix} 0 & 0 & 120 & 0 & 100 & 0 \end{bmatrix}, \quad (29d)$$

$$subst(r_4, a_1, \mathbf{P}, \mathbf{O}) = \begin{bmatrix} 166 & 0 & 0 & 100 & 0 & 0 \end{bmatrix}. \quad (29e)$$

It means that r_1 and r_2 resources may substitute each other, r_3 and r_5 resources are able to substitute each other and r_4 resource is capable to substitute r_1 resource. Other substitutions are not possible. These vectors show that

$$ra(r_2, a_1) = 1, \quad (30)$$
$$ra(r_1, a_2) = 1, \quad (31)$$
$$ra(r_3, a_5) = 1, \quad (32)$$
$$ra(r_5, a_3) = 1 \text{ and} \quad (33)$$
$$ra(r_4, a_1) = 1. \quad (34)$$

Moreover, we suppose that

$$ra(r_i, a_j) = 1, \text{ if } i = j; \quad (35)$$
$$ra(r_2, a_7) = 1 \text{ and} \quad (36)$$
$$ra(r_5, a_8) = 1. \quad (37)$$

This means that a_7 activity can be carried out by r_2 resource, and a_8 activity can be carried out only by r_5 resource. In case of failure, the resources are unavailable as much as shown in *Table 2*.

We suppose that each resource works only in one operation mode and all of the resources have the same error probability. In the example we work only with one product type **(P)**, and we intend to minimize the maximum operation time. Applying the first step of our algorithm, we create the sequence of the resources for each process. This sequence completed with the durations of each activity after user-based modification of the default operation times as illustrated in *Fig.10*.

At this phase we apply the second step of our algorithm and determine the robustness of each processes. In the calculation, we use the following computation method, as we mentioned earlier:

$$t_{max\,end}(proc_i) - t_{start}(proc_i) - TDUR \quad (38)$$

The values obtained are summarized in *Table 3*.

Table 3: The robustness values of the example's processes

process	proc$_1$	proc$_2$	proc$_3$	proc$_4$	proc$_5$	proc$_6$
robustness	16	4	2	0	11	8
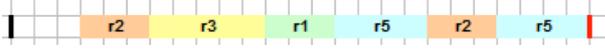
Figure 11: Placement of the second example's basic process onto the time scale (proc$_4$)

Figure 13: Placement of the 2nd example's third process onto the time scale (proc$_2$)

Figure 15: Placement of the second example's fifth process onto the time scale (proc$_5$)

Figure 12: Placement of the second example's second process onto the time scale (proc$_3$)
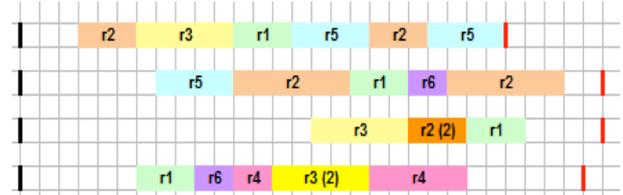
Figure 14: Placement of the second example's fourth process
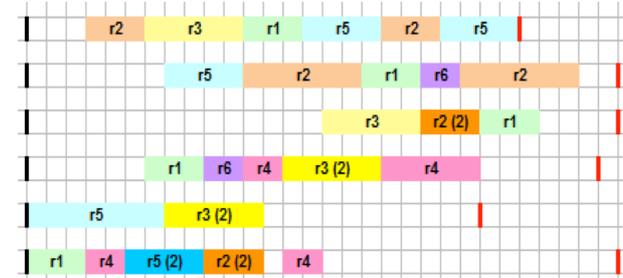
Figure 16: Placement of all processes of the second example onto the time scale (proc$_1$)

Table 3 indicates that the basic process is proc$_4$, which is the least robust process. We place its activities onto the time scale as a chart illustrates this in *Fig.11*. The black line represents the 0 time point and the red one signs the maximal finish time of the process. The second least robust process is proc$_3$. The placement of its activities onto the time scale can be done easily, without any collision as shown in *Fig.12*. The third least robust process is proc$_2$. During its placement there is one collision: its second activity collides with proc$_4$'s fifth activity. However, it doesn't cause any problem, because there are two pieces of resource r_2. The second resource of r_2 resource type can carry out the activity in the originally planned time. The result is shown in *Fig.13*.

Looking for the next least robust process, proc$_6$ is the next. Its fourth activity collides, thus the second piece of r_3 resource type has to be used for carrying it out. Moreover, its final activity also collides. Unfortunately, there is only one piece of r_1 resource type. However, r_2 resource type can substitute an r_1 resource. The problem is that both r_2 resources are occupied at the desired time. There is another resource type, r_4, which is able to substitute an r_1 resource with a bit more necessary time than r_2. As r_4 is free at the desired time, it will take over the task. The time it requires for its task is 1.66 times more than an original r_1-type resource would need for that. *Fig.14* illustrates the results.

The second most robust process is proc$_5$. During the placement of its activities onto the time scale, we find that the second activity collides with proc$_4$'s second activity. Fortunately, there is a second piece of r_3

resource type which can be used freely for the desired time interval. It solves the problem and results a chart as seen in *Fig.15*.

At last, three problematic cases happen during process 1: the collision of the third and fourth activities can be solved by the second piece of the desired resource types; however, in case of the final activity only time shift can solve the collision, because there is only one piece of r_4 resource type and it can not be substituted by any other resource. The result is shown in *Fig.16*.

After placing all activities onto the time scale and solving all collisions, only the fault-tolerance-related improvement need to be done, as the final step of our algorithm. This phase intends to shift the activities of the processes in time, starting from their end by the duration of their preceding activity plus its F value. Because of the maximal finish time constraint, proc$_4$ and proc$_3$ (the first two processes on *Fig.16*) cannot be modified. If we look at *Fig.16*'s third process, its last activity requires $F(r_1) = 2$ time units plus its normal operation time in case of failure. If we take into account the maximal finish time of the process, its place cannot be modified. This is the case for the fourth process, as well. The fifth process in in *Fig.16* ends with r_3 resource. In case of a failure, it requires 8 time units until it reaches the maximal finish time of the process. Taking this into account, only 3 time units remain,
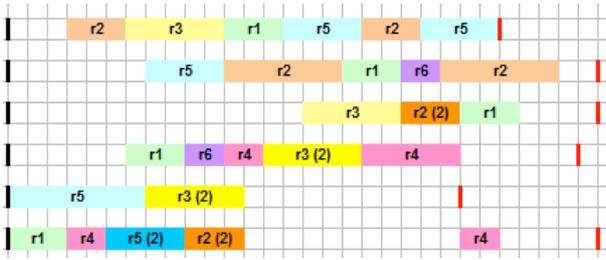
Figure 17: Building fault tolerance into the last process of the second example

Figure 18: Building fault tolerance into the last process of the second example

which is not enough for the previous activity. That's why we do not modify this process, either. The final activity of the last process, $proc_1$ is carried out by an r_4 resource. In case of a failure, it requires 4 extra time units. In order to avoid collision we shift this activity a bit more forward as shown in *Fig.17*. As $F(r_2) = 1$, and a_2 activity requires 3 time units to be carried out by r_2 resource, $proc_1$'s fourth activity can also be shifted in time. Its previous activity would require 6 time units after itself. Its shift would result 1 extra time unit for previous activity, which is not enough for making it ready for tolerating fault. This hinders building in more fault tolerance. The final result of the algorithm is shown in *Fig.18*, in which we have a schedule, which took into account the possible substitutions, the capabilities of the resources; moreover in some places, it tolerates faults of resources without the need of rescheduling.

Conclusion

We presented a model and algorithm for creating schedules, which tolerate some resource-failures and utilize the substitutability possibilities of the resources. Our goal was to establish a model, which can be the basis of applications in different segments of scheduling problems and makes it possible to generate schedules with the optimization criteria related to the operation time.

We introduced an algorithm, which creates an initial schedule on the basis of the input processes and the substitutability of the resources. This can be improved with preparation for likely failures of resources. The operation of the algorithm was presented on two hypothetical case studies.

Acknowledgements

REFERENCES

[1] PINEDO M.L.: Scheduling: Theory, Algorithms, and Systems, 4th ed., Springer, New York, USA, *2012*

[2] BRUCKER P.: Scheduling algorithms, 5th ed., Springer, Osnabrück, Germany, *2007*

[3] BRAHIMI B., AUBRUN C., RONDEAU E.: Modelling and Simulation of Scheduling Policies Implemented in Ethernet Switch by Using Coloured Petri Nets, Proc. 11th IEEE Int. Conf. Emerging Technologies and Factory Automation, Prague, Czech Republic, *2006*, 667–674

[4] BARBA I., VALLE C. DEL: A Job-Shop Scheduling Model of Software Development Planning for Constraint-based Local Search, Int. Journal of Software Engineering and Its Applications, *2010*, 4(4), 1–16

[5] XU J., LIU C., ZHAO X., YONGCAREON S.: Business process scheduling with resource availability constraints, Proc. OTM'10, Hersonissos, Crete, Greece, *2010*, 1, 419–427

[6] SULE D.R.: Production Planning and Industrial Scheduling: Examples, Case Studies and Applications, 2nd Ed., CRC Press, Ruston, LA, USA, *2007*

[7] MURTHY S., AKKIRAJU R., RACHLIN J., WU F.: Agent-based cooperative scheduling, Proc. AAAI Workshop on Constraints and Agents, Providence, RI, USA, *1997*, 112–117

[8] DULAI T., WERNER-STARK Á.: Immediate event-aware routing based on cooperative agents, Proc. Factory Automation, Veszprém, Hungary, *2012*, 144–148

[9] PALOMBARINI J., MARTINEZ E.: SmartGantt – An intelligent system for real time rescheduling based on relational reinforcement learning, Expert systems with Applications: An International Journal, *2012*, 38(11), 10251–10268

[10] NANDANWAR J., SHRAWANKAR U.: An adaptive real time task scheduler, IJCSI International Journal of Computer Science Issues, *2012*, 9(6/1), 335–340

[11] GREGG C., BOYER M., HAZELWOOD K., SKADRON K.: Dynamic heterogeneous scheduling decisions using historical runtime data, Proc. 2nd Workshop on Applications for Multi- and Many-Core Processors, San Jose, CA, *2011*

ESTIMATION OF PARAMETERS FOR AN EXTREMELY LOW FUEL CONSUMPTION INTERNAL COMBUSTION ENGINE-BASED MEGAMETER-III VEHICLE

István Pintér[1][✉] and Mihály Bagány[2]

[1] Department of Informatics, Kecskemét College, Izsáki str 10, Kecskemét, H-6000, HUNGARY
[2] Department of Natural Sciences and Engineering, Kecskemét College, Izsáki str 10, Kecskemét, H-6000, HUNGARY
[✉]E-mail: pinter.istvan@gamf.kefo.hu

Megameter-III is a special vehicle designed and constructed for international competitions by our team. A crucial part of the vehicle is the internal combustion engine, which ought to have extremely low fuel consumption. The experiences of previous competitions confirmed that it is necessary to develop a professional telemetry system, which allows tracking and logging online important parameters of the engine and the vehicle. In this paper, we discuss results for estimating of several parameters of the engine (angular velocity, angular acceleration, the Coulomb-Morin external friction torque and the damping constant). In our work, injector-pulses and inductive sensor's signals were monitored.

Keywords: internal combustion engine, extremely low fuel consumption, angular velocity, angular acceleration, external friction torque, damping constant

Introduction

Since 2010, the racing team at Kecskemét College participates in international competitions of fuel-efficient vehicles. The vehicles were named after 1 Megameter being 1000 km. The first result with Megameter-I in gasoline category was 1588 km L^{-1} (2010, Lausitz, Germany), and the best result of Megameter-II was 2661 km L^{-1} (2011, Nokia, Finland). Megameter-III achieved 2696 km L^{-1} (2012, Rotterdam, Nederland. The team-photo can be seen in *Fig.1*). The latest model, the Megameter-IV has exceeded the 3000 km/litre dream-limit (3082 km L^{-1}; 2013, Nokia, Finland) [1].

Figure 1: Student team photo with the Megameter-III vehicle (Rotterdam, 2010, second place)

During the development of the Megameter series, it became clear that in order to improve the technical parameters of the vehicle it is necessary to establish a telemetry system suitable for measuring, displaying and logging the engine's and vehicle's parameters during the race. This paper presents our recent results in this work, which have already been utilized in developing and construction of the Megameter-IV.

The measurements were taken using the Megameter-III engine, which has the following characteristics:

- one cylinder, four-stroke, air-cooled
- overhead controlled, overhead valve, intake manifold injection
- bore/stroke: 31.5/45 mm
- displacement volume: 45 cm^3
- compression ratio: 14
- power: 700 W
- torque: 2.2 N·m
- specific fuel consumption 235 g kWh^{-1}

There are different application areas and methods for estimation of engines' angular velocity. These estimation procedures are based mainly on crankshaft's rotation angle [2, 3, 4]. In the problem of model-based estimation of mechanical losses, the dependence of instantaneous angular velocity on angular position plays an important role [4]. A second-order spline-interpolation-based method was proposed in [2] for real-time measurement of angular velocity and angular acceleration. Our telemetry system is related to this latter work, and our results proved to be useful in further developing our special vehicle.

The physical model with its approximations and the parameter estimation

In order to achieve ever decreasing fuel consumption it is necessary to uncover the sources of energy losses. For this purpose a suitable physical model of internal combustion engine (ICE) is needed [5]. Moreover, with an appropriate physical model in hand, one can design the measurement for estimating the parameters of the physical model.

Physical model for estimating the external friction torque and damping constant

The instantaneous angular velocity of the crankshaft is changing even during only one crankshaft revolution. It is smaller on the compression stroke and greater in the midst of power stroke. The friction torque affecting the engine, averaged for two crankshaft revolutions depends on the instantaneous angular velocity. By expanding this function up to the second order we get *Eq.(1)*:

$$-M \approx M_0 + \frac{M_1}{\omega_1}\omega + \frac{M_2}{\omega_2^2}\omega^2 \qquad (1)$$

where the indexed quantities are constant and positive, and they have physical meanings as follows: M_0 is the Coulomb-Morin external friction torque, which is independent of the relative speed of frictional surfaces (in this case the relative speed is the angular velocity); $M_1/\omega_1\cdot\omega$ defines the damping constant (the second term corresponds to torque resulting from Newton's internal friction's force, it is proportional to the first power of angular velocity); $M_2/\omega_2^2\cdot\omega^2$ stands for the drag torque characterizing the turbulent flows in pipes; it is in proportion to the square of angular velocity and it is resulting from the flow resistance in the intake and exhausting channels, in the valve-cage and in the air-pipe of the crankcase.

Our task is the estimation of the external torque from the measured data; therefore it is necessary to estimate the angular velocity. The basic equation of rotational motion is shown in *Eq.(2)*:

$$-M = -\Theta\frac{d\omega}{dt} \qquad (2)$$

where after substitution of *Eq.(2)* into *Eq.(1)* we get

$$M_0 + \frac{M_1}{\omega_1}\omega + \frac{M_2}{\omega_2^2}\omega^2 = -\Theta\frac{d\omega}{dt} \qquad (3)$$

Let's assume that the initial value of the angular velocity is ω_0. By separating the variables and integrating we get *Eq.(4)*:

$$\int_{\omega_0}^{\omega} \frac{d\omega}{M_0 + \frac{M_1}{\omega_1}\omega + \frac{M_2}{\omega_2^2}\omega^2} = -\int_0^t \frac{dt}{\Theta} . \qquad (4)$$

The result of the integration is shown in *Eq.(5)*.

$$t = -\frac{2\Theta}{\sqrt{4ac-b^2}}\,\mathrm{atan}\left(\frac{2c\omega+b}{\sqrt{4ac-b^2}}\right) + \\ + \frac{2\Theta}{\sqrt{4ac-b^2}}\,\mathrm{atan}\left(\frac{2c\omega_0+b}{\sqrt{4ac-b^2}}\right) \qquad , \qquad (5)$$

where three new variables were introduced for more readability as shown in *Eq.(6)*.

$$a = M_0 , \quad b = \frac{M_1}{\omega_1} , \quad c = \frac{M_2^2}{\omega_2^2} . \qquad (6)$$

By solving *Eq.(5)* for ω one can get the final solution. However, instead of doing so, we analysed two approximate solutions of *Eq.(4)*.

As we will see below, these approximate solutions fit real measurements well. In order to estimate the engine's friction torque and damping constant long measurement duration (e.g. 20 s) is necessary. The relevant physical quantities can be estimated from the measured injector pulses.

Approximation 1: the friction torque is independent from angular velocity

By assuming that the other two terms are negligible compared to external friction M_0 the differential *Eq.(4)* simplifies to:

$$\int_{\omega_0}^{\omega} \frac{d\omega}{M_0} \approx -\int_0^t \frac{dt}{\Theta} , \qquad (7)$$

and its solution is:

$$\omega = \omega_0 - \frac{M_0}{\Theta}t , \qquad (8)$$

where ω_0 is the value of angular velocity at the beginning of deceleration.

The angular velocity of the engine's crankshaft has to be measured as the function of time for parameter estimation. Namely, after line-fitting to measured data the intercept and slope can be computed from relationship $\hat{\omega} = A - Bt$, so both the estimation of Coulomb-Morin external friction torque and the estimation of initial value of angular velocity can be determined: $\omega_0 = A$, $M_0 = B\Theta$.

Approximation 2: the friction torque is independent from the squared value of angular velocity

When the engine is running with wide open throttle, and the friction torque of pipe-flow losses is negligible comparing to external and internal losses, the differential equation to be solved is the following:

$$\int_{\omega_0}^{\omega} \frac{d\omega}{M_0 + \frac{M_1}{\omega_1}\omega} \approx -\int_0^t \frac{dt}{\Theta} \; . \tag{9}$$

The solution of *Eq.(9)* is:

$$\omega = \left(\omega_1 \frac{M_0}{M_1} + \omega_0\right) e^{-\frac{t}{\tau}} - \omega_1 \frac{M_0}{M_1} \; , \tag{10}$$

where $\tau = \omega_1\Theta/M_1$ is the time constant characterising the angular deceleration. By measuring the engine's crankshaft's angular velocity as the function of time, and after fitting a biased exponential-type curve we get the estimation $\hat{\omega} = Ae^{-Bt} - C$, from which the estimates of physical quantities are the following:
- the external friction torque is: $M_0 = BC\Theta$,
- the damping constant is: $M_1/\omega_1 = B\Theta$,
- the initial value of angular velocity resulted from curve-fitting is: $\omega_0 = A-C$.

Estimation of the instantaneous torque and the running irregularity

The instantaneous angular acceleration as the derivative of the instantaneous angular velocity is needed for estimation of the instantaneous torque. Moreover, knowing the instantaneous angular velocity, another important parameter of the engine, the running irregularity can be determined. From the point of view of measurement, the estimation of instantaneous angular velocity requires the duration of two crankshaft revolutions (e.g. 100 ms). For measuring this, an inductive sensor was used.

When determining the running irregularity, the engine and the dynamometer were disconnected and the flywheel ring gear teeth together with an inductive sensor served as signal source. The induced voltage from the sensor can be approximated as:

$$\frac{U}{U_0} = \frac{(\omega + \beta t)}{\omega_0}\cos\left(z\omega t + z\frac{\beta}{2}t^2\right), \tag{11}$$

where U is the output signal of the sensor, U_0 is the amplitude of the sensor corresponding to the constant angular velocity of ω_0, ω is the angular velocity of the crankshaft, β is the angular acceleration of the latter, z is the number of teeth of ring gear.

In case of zero or small values of angular velocity the relationship *Eq.(11)* simplifies to
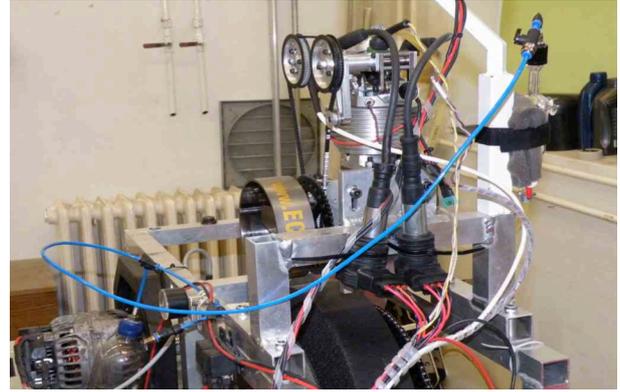
Figure 2: Arrangement for measuring the friction torque

$$\frac{U}{U_0} = \frac{\omega}{\omega_0}\cos(z\omega t) \; . \tag{12}$$

By assuming that during the pass-time of one tooth (tooth-time) the angular velocity is not changing significantly, the signal amplitude during the tooth-time can be well approximated by a constant. That is, the approximated sensor signal when passing the zth tooth is given as $U/U_{0k} = \cos(z\,\omega_k\,t)$. The running irregularity is defined as a ratio in *Eq.(13)*:

$$\delta = \frac{\omega_{max} - \omega_{min}}{\omega_{average}} \tag{13}$$

Measurements, estimations and evaluations

The measurements were accomplished in the Student's Workshop of GAMF Faculty using an Agilent DSO-X 2002A digital storage oscilloscope. The oscilloscope is two-channel type with upper input cut-off frequency of 70 MHz, maximum sampling frequency of 2 GS/s and with sample storage of 50000 samples.

The algorithms used for evaluation of measured data were intentionally different. In one case the data series were not re-sampled and the differentiation was approximated by finite differences, and in the other case both re-sampling and smoothing derivative algorithm were applied. Moreover, the computation method of the trend of angular acceleration was also different.

Estimation of the external friction torque and damping constant

In this setup, an extra flywheel was driven by the engine fixed on the test stand. The measurement setup can be seen on *Fig.2*. The reduced resultant moment of inertia of synchronously rotating components with the crankshaft was $\Theta \approx 0.0534$ kg m^2.

The dynamometer's friction torque was approximately constant with a value of $M_{dynamometer} \approx 0.2$ N m.
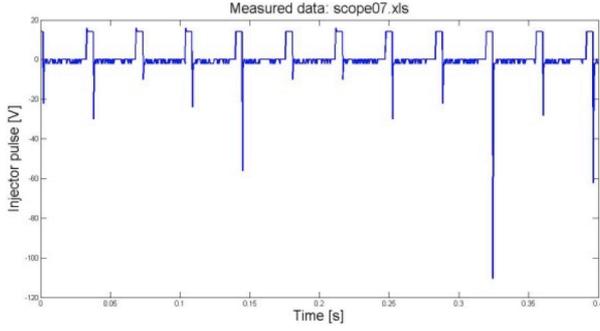
Figure 3: Measured injector pulses using a power MOSFET

Figure 4: Linear model: the brake torque is independent of angular velocity

Table 1: Results of the fitting of data shown in *Fig.4*

Resampling	A, rad s^{-1}	B, rad s^{-2}	R^2
No	353.78	12.103	0.9954
Yes	346.75	10.952	0.9981

Table 2: Estimation of physical quantities

Resampling	n_0, rpm	M_0, N·m	M_{ICE}, N·m
No	3378	0.650	0.450
Yes	3311	0.585	0.385

Figure 5: Biased exponential model: the braking torque is independent of the square of angular velocity

The engine's injector was triggered by a signal from the crankshaft, one injection occurred in every two crankshaft revolutions. The oscilloscope was connected to the injector's solenoid, and the solenoid's signal was sampled with a sampling frequency of 2.5 kHz, that is the sampling interval was T_s = 400 μs. The part of measured injector pulse-signal can be seen on *Fig.3*. The crankshaft's angular velocity ω = ω (t) as a function of time was determined from the stored samples, with duration of 20 seconds. At the beginning of the measurement, the engine was accelerated to 4000–4200 rpm, then the injector was disconnected from the intake manifold, and the engine was allowed to decelerate freely. By defining a suitable trigger level, the elapsed time T between two consecutive rising edges of the impulse can be determined. The instantaneous frequency is two times the reciprocal value of T ($f_{instantaneous} = 2 / T$), because every two revolution results in one impulse. The instantaneous angular velocity can be computed from this instantaneous frequency value. As the resulting angular velocity value sequence is non-equidistant, it has been re-sampled using cubic spline interpolation. First, a linear model was fitted *(Fig.4)*. The results are summarised in *Table 1*. The two methods gave similar parameters, but the resampling resulted in somewhat better goodness-of-fit. The estimation of the physical quantities from parameters shown in *Table 1* can be seen in *Table 2*. The estimations resulted by the two methods are similar in case of model of *Eq.(7)*.

In case of model of *Eq.(9)*, the fitting has been fulfilled on the resampled data series. The results are demonstrated on *Fig.5*.

From the above results the external friction torque is $M_0 = B·C·\Theta \approx 0.419$ N·m; the average braking torque of the engine is $M_{engine} = M_0 - M_{dynamometer} \approx 0.219$ N·m; the damping constant is $M_1/\omega_1 = B\Theta = 6.94·10^{-4}$ N·m·s. The initial angular velocity and revolution are $\omega_0 = A − C \approx$ 352 rad s^{-1}, n_0 = 3358 rpm.

We can conclude that the two types of models gave similar results, but the biased exponential model fitted somewhat better the measured data than the linear model, thus the former are preferred. Furthermore, the presented data indicate that both the average breaking torque of the engine and the damping constant are acceptable small, therefore the engine is acceptable from this point of view.

Estimation of the instantaneous torque and the running irregularity from measurements

When the running irregularity was measured, the engine and the dynamometer were disconnected. The moment of inertia of the parts running synchronously with the crankshaft was Θ = 4000 kg·mm^2. A ring gear fixed to the crankshaft was used as signal source (number of teeth is z = 66) together with an inductive sensor fitted from the teeth in a distance of 4 mm. An oscilloscope was connected to the sensor and the voltage signal was sampled with 500 kHz sampling frequency *(Fig.6)*. It is necessary to use such a great sampling rate because of the following. In order to acceptably estimate the instantaneous angular velocity, it is necessary to measure N samples during the tooth-time. In case of nominal revolution of n_0, the sampling frequency is f_S =

Figure 6: Setup for measuring the instantaneous torque
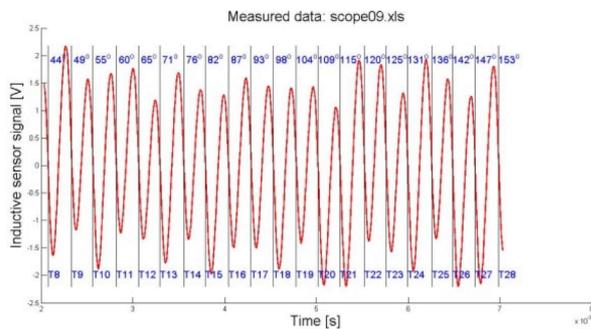
Figure 8: Inductive sensor's signal labelled with the crankshaft's angle and tooth number (angles are in degrees)

Figure 7: The raw and smoothed signal of inductive sensor in case of SAVITZKY-GOLAY smoothing

Figure 9: Instantaneous angular velocity values and the FFT-based trend

Figure 10: Estimation of instantaneous acceleration as smoothing derivative of the trend of instantaneous angular velocity

$n_0 \cdot N \cdot z$ / 60, that is in case of $n_0 = 4000$ rpm and $N = 100$ samples, $f_S = 440000$ Hz. The nearest possible setting of the DSO is 500 kHz. The capacity of the sample storage is 50000 samples, so the duration of the measurable signal is 0.1 s. As it is greater than the duration T_2 of two crankshaft revolution, the measurement of at least one full engine cycle is feasible using this DSO ($T_2 = 2 \cdot (60/n_0) = 0.03$ s). The instantaneous angular velocity is the reciprocal of the time between two falling edge of the sensor's signal, so the raw signal had to be smoothed before the zero-crossing detection.

Two different smoothing methods were implemented in this case as well: a 10-point moving average smoothing and SAVITZKY-GOLAY smoothing with tenth order polynomial and 31-point length window (for the latter see *Fig.7*). Because the number of teeth is known, the exact value of the crankshaft angle also with the tooth number can be tracked (see *Fig.8*). As the angular velocity values are non-equidistant, cubic spline interpolation-based re-sampling was applied as in the previous section.

Due to the fast sampling rate, we can track very short-time changes in angular velocity values during 100 ms. However, our task is to estimate the crankshaft's instantaneous angular velocity. For this purpose the trend of the angular velocity curve estimated at teeth is necessary. The trend was determined on the one hand by using a 10-point moving average filter, on the other with FFT-based filtering (for

the latter see *Fig.9*). In the first case the time of averaging was around 2 ms. Using this value for the time-constant τ of a low-pass filter for the FFT-based computation, for the cut-off frequency of the low-pass filter we get $f_H = 1$ / $(2 \cdot \pi \cdot t) \approx 80$ Hz. Following its definition, the running irregularity was computed using the maximum, minimum and average values of the angular velocity trend shown in *Eq.(14)*.

$$\delta = \frac{\omega_{max} - \omega_{min}}{\omega_{average}} = \frac{389.3 - 379.7}{384.7} \approx 0.02 \ . \qquad (14)$$

The estimation of instantaneous acceleration was computed with differences in case of moving average-based trend. In case of FFT-based trend, the SAVITZKY-GOLAY smoothing derivative algorithm was applied (tenth order polynomial, 31-point length window).

Fig.10 illustrates the instantaneous acceleration as a function of time. In this measurement the maximum

value of angular acceleration was 1113 rad s^{-2}, so for the corresponding instantaneous torque we get:

$$M_{max} = \Theta\beta_{max} \approx 0.004 \text{ kg} \cdot \text{m}^2 \cdot 1113 \frac{\text{rad}}{\text{s}^2} \approx 4.5 \text{N} \cdot \text{m}. \quad (15)$$

Nearby the power stroke instant higher value of angular acceleration can be estimated. In that measurement the corresponding instantaneous torque was of 16 N·m (for details see Ref. [6]). The crankshaft's torque gets to wheel by a drive train with a reduction ratio of 7, so the maximal torque at the wheel is 112 N·m. The wheel's maximal slipless torque transmission is 40–50 N·m that is during the maximal instantaneous acceleration the wheel will slip on the road, which results in energy loss. This effect ought to be concerned in designing the drive chain of Megameter-IV, e.g. torque-damping torsion clutch should be built in.

Conclusions

Two methods were elaborated for instantaneous speed and acceleration of an internal combustion engine via on-line measurement. Both the injector-pulse-based and the inductive sensor-based methods gave useful data for developing the next engine, namely the Megameter-IV. The results presented in this paper are the numerical estimation of the engine's internal friction, the damping constant, the instantaneous angular velocity, the instantaneous angular acceleration and the estimation of the instantaneous torque of the engine during the power stroke. These results provided the foundation for the design and construction of the on-board electronic control unit, which can be seen on *Fig.11*.

Acknowledgement
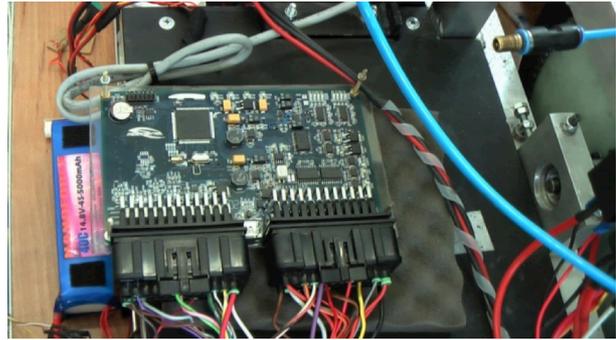
Figure 11: ECU is under development for motor control and telemetry

REFERENCES

[1] www.eco-marathon.eu (last accessed: May 20, 2014)

[2] BÉLANGER P.R.: Estimation of angular velocity and acceleration from shaft encoder measurements, Proc. IEEE International Conference on Robotics and Automation, Nice, France, *1992*, 585–589

[3] STOTSKY A., FORGO A.: Recursive spline interpolation method for real time engine control applications, Control Engineering Practice, *2004*, 12, 409–416

[4] CRUZ-PERAGÓN F., PALOMAR J.M., DÍAZ F.A., JIMÉNEZ-ESPADAFOR F.J.: Fast on-line identification of instantaneous mechanical losses in internal combustion engines, Mechanical Systems and Signal Processing, *2010*, 24, 267–280

[5] BAGÁNY M.: Internal combustion engines, Kecskemét College, GAMF Faculty, *2011*, 192 (in Hungarian)

[6] PINTÉR I., BAGÁNY M.: Algorithm development for the Megameter-III telemetry system to estimate the parameters of extremely low fuel consumption internal combustion engine, Proc. Factory Automation, Veszprém, Hungary, *2013*, 21–22, 94–99

INFLUENCE OF CAN FLATNESS ON HEAT DISSIPATION OF ALUMINIUM ELECTROLYTIC CAPACITOR

László Kovács[1✉], László Gál[1], and Dénes Fodor[2]

[1] Department of Aluminium Electrolytic Capacitor Development (SZ ALU PD), Epcos LLC, Szt. László Str. 6, Szombathely, H-9700 HUNGARY
[2] Automotive Mechatronics Department, Institute of Mechanical Engineering, Faculty of Engineering, University of Pannonia, Egyetem Str. 10, Veszprem, H-8200 HUNGARY
✉E-mail: laszlo.kovacs@epcos.com

The lifetime of aluminium electrolytic capacitors highly depends on their core temperature. Heat dissipation in general applications happens by the extended cathode, which is in contact with the inner side of the can. In the case of heat sink applications, the most important heat transfer phenomenon is the heat conduction through the bottom of the aluminium can. The quantity of the dissipated heat is in direct proportion to the size of the heat transfer surface. The more dissipated heat may increase the lifetime of the capacitor. Therefore, the flatness value of the can bottom is critical. This paper presents a flatness measurement method, which can successfully replace the equipment for a more complex and more expensive 3D measurement. It discusses an implementation of a measurement environment, where data acquisition and visualization are automated by a LabVIEW-based software. In addition, this study deals briefly with the influence of production processes on the flatness value of the capacitor produced by leading manufacturers.

Keywords: aluminium electrolytic capacitor, measurement automation, heat dissipation, graphical programming environment, flatness measurement

Introduction

The aluminium electrolytic capacitor is the most commonly applied capacitor type, due to the fact that the capacitance and voltage range of the components is wide. The rated voltages change from 5 V to 550 V, while the capacitance 1 μF to 3 F. These are applied in many fields of industry, such as energetics, power electronics, automotive application, etc. and used for energy storage, smoothing and filtering function. The lifetime of the capacitor highly depends on its core temperature. Each decrease of 10 ºC doubles the lifetime [1]. Operating temperature is determined by the ambient temperature, by the applied ripple current, by the used voltage, and by the equivalent serial resistance (ESR). There are two ways for the reduction of core temperature. The first one is the extended cathode foil and the second one is the cooling of the capacitor can. Cooling in general applications is realized by applying a heat sink at the bottom of a capacitor can. This paper presents the basic construction of capacitors and the most important production steps, introduces the heat conduction between the can and the environment and describes the effect of the flatness of the can to heat conductivity. It also presents the entire measurement environment (measurement station, data acquisition and evaluation software), the database behind the measurement system that stores the results. At last, it discusses the effect of production processes to the measured flatness, and presents the flatness of capacitors produced by leading capacitor manufacturer companies.

Structure and construction of aluminium electrolytic capacitor

The winding of an aluminium electrolytic capacitor contains two foils and papers [1]. These are rolled together tightly into a winding. The material of the anode, positive foil is aluminium with purity higher than 99.9%. The foil has been etched [2] to increase the effective surface area (and thus the capacitance of the capacitor). As a result, the effective surface area becomes typically 20–40 times larger than the plain area of the foil. On the etched surface of the foil an aluminiumoxide layer [3] has been generated electrochemically. The forming voltage of the anodized aluminium foil [4] is 30–60% higher than the rated voltage of the capacitor. The material of the cathode foil is also aluminium and itself has a thin oxide film (the forming voltage is only a few volts regardless of the rated voltage). It is typically etched to slightly increase the surface area. The anode and cathode foils are connected to aluminium tabs, which are coming out from the winding and are riveted to the aluminium terminals of the cover disk. The tab foils are not etched but are provided with an oxide layer made by

electrochemical oxidization. Before being housed in a suitable container, the complete winding is impregnated with electrolyte. After housing the edges of the can are curled down. Before being sleeved and packed, capacitors are first aged. The purpose of this stage is to repair any damage in the oxide layer and thus to reduce the leakage current to very low levels. During manufacturing there are two processes (curling and aging processes), which apply mechanical stress to the casing of capacitors.

The effect of flatness of aluminium cans on heat conductivity

In common applications with usage of extended cathode (the cathode is wider than the anode and exposed to the bottom of the winding) or different kinds of capacitor cooling (for example air or water cooling) are applied to reduce the core temperature that effects its lifetime. Earlier investigations showed that there are three types of heat transfer phenomenon between capacitor and the environment [5]: (i) heat conductivity (mostly at the bottom), (ii) radiation (mostly at the side), and (iii) convection. This paper only deals with heat transfer from the bottom of the capacitor to the environment. From the general theory of thermodynamics the heat conduction can be described by *Eq.(1)*.

$$P = kA \frac{dT}{dx} , \qquad (1)$$

where P represents power of the heat flow (W); A is the size of the tangential surface (m^2); dT/dx is the temperature gradient (K m^{-1}); and k stands for heat conductivity of the material (W Km^{-1}). In our case the linear approximation is acceptable. As already described, the rate of the heat flow depends on the size of the tangential plane.

Definition of the flatness was adapted from Ref. [6]. It is given as the distance between the two closest tangential planes of the bottom of the aluminium electrolytic capacitor can. In the case of capacitors with higher flatness the tangential surface of the aluminium can is lower. The distance between the bottom of the capacitor and the heat sink is filled by air. It is well known that the heat conduction coefficient of the air is much lower than the aluminium. The aspect ratio between the two constants is in the order of magnitude ~10^4. Therefore the flatness of the bottom in the case of regular heat sink applications is an important value.

Theoretical background of used measurement method

There are several different methods for measuring the flatness of a surface [4]:

1. The entire surface is compared with a known reference surface (holistic methods);
2. Points on the surface are compared to a reference plane;
3. Straightness of the lines in the surface is measured.

The holistic methods (e.g. holographic and interferometric methods) are not suitable for large surfaces with more than 90 mm diameter of screw terminal capacitor [7]. The third method is too complex and time-consuming for applying in serial production. Therefore, a point-to-point method (second type) was chosen for this investigation. Of course, the expensive optical methods may be more precise (in the order of 100 nm deviations) [8], but in our case this high precision is unnecessary. On the other hand, the point-to-point method is very fast, can be used easier and cost-efficient. For point-by-point methods, the grid size is a critical issue, which determines the lowest limit for observable surface irregularities. In our case, approximately 60 cm^2 surface area needs to be characterized. The height difference was measured between N points within the bottom. Trials were made to find the applicable grid. Finally, a 9-point-grid was chosen with circular symmetry, because the measurement time is short enough and the accuracy of the measurement is acceptable.

Least mean squares (LMS) method was used for plane fitting, because the randomly distributed measurement errors have the smallest influence to the results [9]. *Eq.(2)* was used for the LMS:

$$\begin{bmatrix} \sum_{i=1}^{N} x_i^2 & \sum_{i=1}^{N} x_i y_i & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i y_i & \sum_{i=1}^{N} y_i^2 & \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} y_i & \sum_{i=1}^{N} 1 \end{bmatrix} \times \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} z_i x_i \\ \sum_{i=1}^{N} z_i y_i \\ \sum_{i=1}^{N} z_i \end{bmatrix} , \qquad (2)$$

where N is the number of the grid points; x_i, y_i are coordinates within the measured surface [mm]; z_i corresponds to the measured height compared to z_1 of the ith point (mm), a, b, c are constants for the fitting.

The fitted plane is described by *Eq.(3)*.

$$z(x,y) = ax + by + c \qquad (3)$$

By solving the system of linear equations, for example with Gauss-elimination the fitted plane is given. The Gauss-elimination is not necessary if the grid points are centrally symmetric. In this case, the input matrix is diagonal, so the solution of the equation system is simple. The distance between the individual points and the fitted plane was calculated. The sign of this distance describes the location of the point respect to the fitted plane. The flatness is the distance between the farthest points above and under the fitted plane. The *Fig.1* shows an example of measured points and the fitted plane.
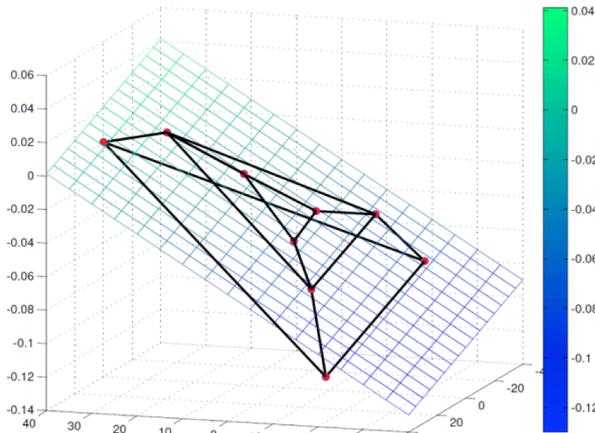
Figure 1: Example of the measured points and the fitted plane

Figure 2: Graphical user interface of the data acquisition software

Figure 3: Graphical user interface of the evaluation software
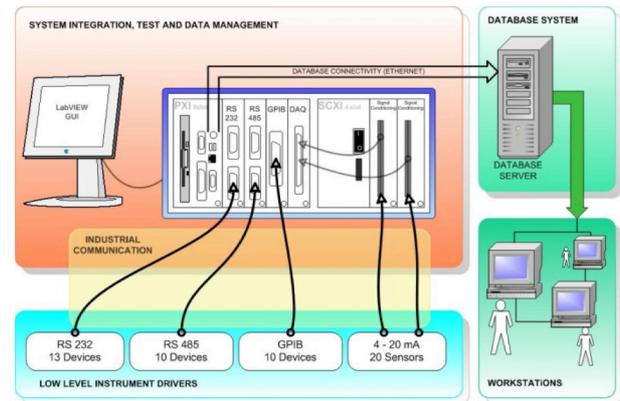
Figure 4: Architecture of the electrolyte measurements

Implemented measurement environment

The base of the hardware was a measuring pad made from stainless steel, which has an extremely smooth surface. By the LMS method, the error that comes from the tilt of the steel pad was avoided. To stay clear of the effect of the temperature gradient, the measuring equipment was kept under controlled environmental conditions. The data acquisition programme for the flatness measurement was written in LabVIEW with contains two major parts. The first one collects the data from the controlled hardware (height meter) and stores them into a local database in MS Excel format. This software communicates through an RS-232 port with the measuring equipment. The user sets the actual position, pushes a button and the module registers the results. The graphical user interface developed is shown in *Fig.2* with Hungarian annotations. The software displays the number of actual measurement points and controls the entire process. For example, if the flatness value exceeds a predetermined limit, the software sends a warning to the operator.

The second module can extract the results from the database of an examined capacitor(s) or the user can import the measured values from Excel. If the results are in the input table, the evaluation module can calculate the flatness of the capacitor can. The graphical user interface of the second module can be seen in *Fig.3*.

The implemented automation environment fits into the measurement automation system (MAS) [10] being used at the Aluminium Electrolytic Capacitor Development Department of Epcos LLC. The system uses two frameworks: a LabVIEW based framework for data acquisition and an ASP.NET based framework for data management.

The measurement part contains automated data acquisition measurements, which are connected to the electrolyte (conductivity, pH, viscosity, etc.) and different kind of capacitor tests. (Lifetime, surge voltage and storage test, etc.). The experiments on the electrolyte are controlled by an NI-PXI, which is connected to the database (*Fig.4*). PXI stand for PCI eXtensions for Instrumentation. These platforms are used as a basis for building of electronic test equipment, automation system, etc. PXI chassis can handle many modules for example plug-in data acquisition (DAQ) cards, communication cards like RS-232, and different analogue and digital I/O boards. The capacitor measurements are controlled by computers and not by the NI-PXI, because the current implementation measures the low and high voltages separately.

The ASP.NET part of the system was developed for data management and evaluation. This is a software module that contains useful tools that facilitate data handling. Data management module simplifies the registration of the constructive properties of the capacitor (like anode foil, cathode foil, type of can, cover disk, etc.) and helps the data storage of applied
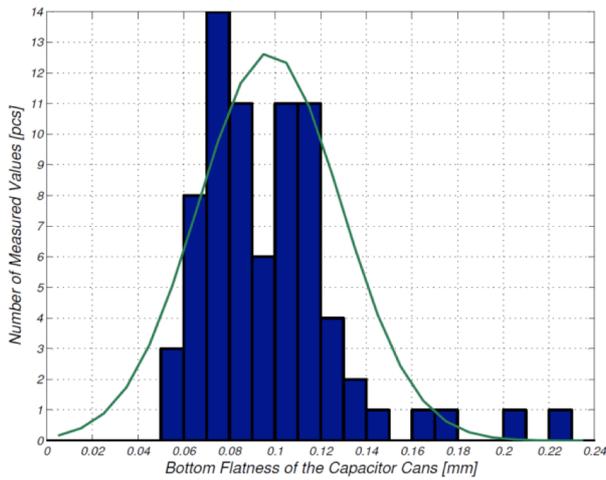
Figure 5: The measurement results from Incoming Inspection Department

Table 1: Results of the ripple current measurement

Average Flatness (mm)	Average of Temperature at equilibrium state (°C)
0.102	90.6
0.259	90.9
0.509	91.5

voltage and current, ambient temperature, etc. The evaluation part supports the evaluation process by generating a standardized report. The user can tailor the reports according to its needs by the Report Generation tool. The desired data appear in a representative way and even the trends of the parameters can be shown.

The goal of the MAS is to automate the previously manual measurements and eliminate paper-based registration. There are many advantages like making the measurements more precise, more reliable and fault tolerant, running multiple measurements in parallel, which all contribute to speed up the research and development of new component and devices.

Change of flatness value during the manufacturing

The flatness of the capacitor can is measured at the incoming inspection and at the final measurement of the production. The initial value of the flatness of the capacitor is determined by the incoming aluminium cans. These parts are produced by cold extrusion of aluminium slugs, so the irregularities of the cans is in the order of magnitude ~10 μm, which is negligible for our purposes.

The *Fig.5* shows the flatness values measured at the incoming inspection of the capacitors. It can be seen that the mean of the measured flatness values is 0.09 mm. As mentioned above, the flatness is mainly affected by two procedures: curling and aging. During curling the can is closed hermetically by curling back the edges of the aluminium can. This curled edge sinks into the rubber ring of the cover disk. As a consequence,
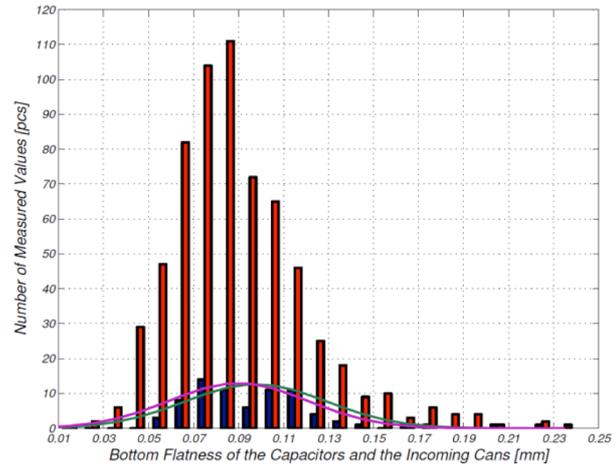
Figure 6: The effect of the production processes to the flatness of capacitor

the cover disk presses down the winding. It can affect the flatness value, since for example concave shaped aluminium can turn into a convex shape. During aging, the flaws of the anode foil are repaired by applying voltage to the capacitor and placing it in an oven. As a side effect gas is generated. The pressure of the generated gas can change the geometry of the can. From the distribution of the flatness values the effect of Epcos production steps (coloured curves in *Fig.6*) can be estimated. The results show that the different production steps do not influence significantly the bottom flatness of the capacitor at the Epcos.

Experimental results regarding to core temperature measurements

Two experiments were completed for investigating the importance of the flatness value of the capacitor can. Sample capacitors were produced with different bottom flatness values. The aluminium cans were made to be convex. These were ordered directly from the supplier and the used parts were sorted out. Regular heat sinks were assembled to the bottom of the parts. The tangential planes of the heat sinks were flat (~2 μm) due to their grinded surface. The heat sinks were fixed to the bottom of the samples with the bottom screw of the capacitors. The mechanical stability of the fixation was made with torque wrench. Also guaranteed the same fixation level in each case. During the tests, forced air-cooling was not used. Heat conductive wires were applied to the core of the samples, which allowed us to measure the core temperature of the capacitors directly. The core temperature measurement system is shown in *Fig.7*.

In case of the first measurement, sample capacitors with very similar equivalent serial resistance (ESR) and leakage current were selected. These parts were taken to 85 °C oven and rated voltage (450 V) was applied to them for 100 h. This preparation is necessary to avoid the different behaviour of the leakage current of the capacitors (isolated from each other to avoid the heat transfer). During the experiment a sinusoidal current

Figure 7: Core temperature measurement of a capacitor

Figure 8: The behaviour of core temperature in case of different bottom flatness values

Table 2: Flatness results for products from leading manufacturers

Company	Capacitor Series	Flatness (mm)
Epcos	B43586	0.095±???
Hitachi	FXR	0.158±???
Kemet	ALS	0.225±???
Nichicon	NT	0.330

was applied at 50 Hz. The current and the temperature (the upper category temperature) of the test are specified by the data sheet of the capacitor. In this case the examined capacitor was B43586A5278Q (C = 2700 μF, UR = 450 V, I (85 °C, 100 Hz) = 12 A), the applied current (multiplying with the frequency factor) was 10.8 A. The selected samples were placed into an oven at 85 °C and a rated voltage of 450 V was applied. The core temperature was measured during the test. The equilibrium temperature was measured and compared. Test was started with the remaining three prepared samples. In case of the second measurement the behaviour of the capacitors was tested against transient heat. The parts were heated up to 85 °C and removed from the oven after 10 hours. The core temperature of the capacitors was measured during cool down time. The recorded core temperature curves *(Fig.8)* show exponential decay. The time-constant of the decay depends on the flatness value of the bottom. As a comparison, the flatness values for screw terminals capacitors from leading manufacturers were inspected. *Table 2* summarizes the flatness results by leading manufacturers.

Conclusion

It is possible to measure the flatness of a capacitor can with the presented and implemented measurement environment. The expensive 3D measurement equipment can be replaced with this cost-effective method. The measurement and evaluation software was implemented in LabVIEW programming environment, which perfectly fits into the previously presented measurement automation system, which is currently being used at the aluminium electrolytic capacitor

product development department of Epcos. The faults of previously manual measurements have been lowered with the help of automation and data acquisition software. This study discussed the basic construction of a capacitor and highlighted the critical construction steps from the aspect of flatness. In addition, it reveals the flatness values of other capacitor manufacturers' products.

Acknowledgments

REFERENCES

[1] TDK Corporation, Aluminium Electrolytic Capacitors, *2013*

[2] PARK D., KIM H.: Electrochemical Etching of Aluminium through Porous Alumina, Analytical sciences, *2001*, 17(supplement), a73–a76

[3] CHURCH H.F.: The dielectric properties of anodic aluminium oxide films, Proc. IEE – Part B: Electronic and Communication Engineering, *1962*, 109, 399–403

[4] CHEN C., MASS W., HUTCHINS G.: Anodization of aluminium electrolyte capacitor foil, U.S. Patent 4,481,084, *1984*

[5] PARLER S.G. JR.: Thermal Modelling of Aluminium electrolytic Capacitor, IEEE Industry Applications Society Conference, *1999*

[6] ISO/TS 12781–1: Geometrical Product Specifications (GPS) – Flatness – Part 1: Vocabulary and parameters of flatness, *2003*

[7] DE GROOT P.J.: Grating interferometer for flatness testing, Optics Letters, *1996*, 21(3), 228–230

[8] MEIJER J.: Accuracy of Surface Plate Measurements – General Purpose Software for Flatness Measurement, Annals of the CIRP, *1990*, 39, 545–548

[9] HAITJEMA H., MEIJER J.: Evaluation of surface plate flatness measurements, European Journal of Mechanical Engineering, *1993*, 38(4), 165–172

[10] KOVÁCS L., FODOR D., KLUG O., ENISZ K.: Measurement Automation System for Aluminium Electrolyte Capacitor Development, IEEE Instrumentation & Measurement Magazine, *2013*, 16(3), 38–43

FINE TUNING OF AUTOMATED ASSEMBLY MACHINES USING VIDEO ANALYSIS

Alíz Katona[1⊠], Viktor Kovács[2], Dóra Tasner[3], and Zoltán Kovács[3]

[1] Transmoduls , József A. u. 9., Veszprém, HUNGARY
[2] Department of Automation and Applied Informatics, Budapest University of Technology and Economics, Magyar Tudósok krt. 2., Budapest, 1117, HUNGARY
[3] Department of Supply Chain Management, University of Pannonia, Egyetem u. 10, Veszprém, 8200, HUNGARY
⊠E-mail: katona.aliz@transmoduls.hu

Human/machine interaction is of still importance in the age of automation. Automated assembling devices are customizable products that need basic settings before installation. Capabilities of operators have to be taken into consideration in this process. This paper presents a procedure based on video recording to determine time data for machine settings. It allows the engineers to forecast the expected capacity of individual workstation and the entire production line. A case study from industry is used to illustrate the application of the system.

Keywords: automated assembly machine, human/machine interaction, method time measurement, Maynard operation sequence technique, video based process analysis

Introduction

Stopwatch, photo and film technologies were used commonly in factories to study how humans work. Analogue video technology had some disadvantages compared to traditional film, which contributed to opening a gap in their applications. The advent of digital age in video technology eliminated most drawbacks. Today videos have numerous benefits over the film and old videos including framing feature, easy to access and wide spread, affordable equipment, no additional personnel or lab needed, and software supported processing is possible. Due to the convergence of the technology, recordings became files that can be acquired by a wide variety of equipment including handheld devices.

We carried out analyses in production factories, observing human work where the machine time (process time: PT) was supposed to be unalterable. There were no reason to observe automates, since they did their job in the pace prescribed by the programmer. This is not the case during the development and production of the automated workstations. Machine times have to be in sync with human capabilities. *Fig.1* shows the layout of analysis in the production project of the customised automated assembly station.

We have carried out analyses at Transmoduls Ltd. in order to help programmers to tune the machines before delivering to the customer. As the first such application of our methodology it consisted of video recordings and software components. The outcome of these analyses are measuring times (time study) or improvement of process (method study). We focused on the former in the given study.

Literature Overview

The interference of operators and their workstations is critical in manufacturing systems [1]. There are many ways to ground process improvement starting from simple observation until the precise measure of data. Motion pictures can be used for time and motion study. The widely accessible digital video technology allows repetitive observations, measuring time and analysing information using software applications. Nowadays, several companies focus on digital time analysis.

Elnekave and Gilad verified the method's abilities for rapid generation of time standards through five case studies. They compared the performance of work analysts for making a Maynard operation sequence technique (MOST) study using the digital video-based approach *versus* using videotape analysis. The result showed that the use of a digital approach may save ~40% of the time devoted to work measurement [2].
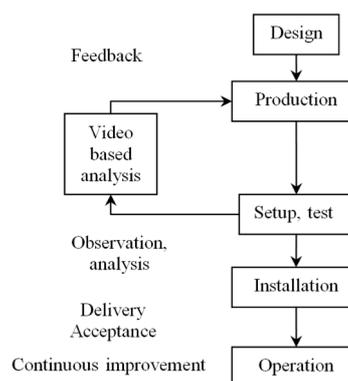
Figure 1: The role of analysis in the project

As an example, the usage of video-based time and motion study is common in health care. A study was conducted to identify ambulance- and operator factors affecting cardiopulmonary resuscitation device in such environment. The aim was to improve the quality of resuscitation in ambulance [3]. Furthermore, a new technique demonstrates the feasibility of tracking workers from statistically placed and dynamically moving cameras [4]. This study reviewed existing techniques to monitor workforce and describe areas, where a work may be improved in engineering application. Typical construction site video is processed and analysed to determine the most appropriate tracking method for the video.

MA *et al.* proposed a new framework on the basis of motion-tracking and digital human simulation technique for motion-time analysis of manual operations. They track the worker's operation during completion of a task. Then the motion data are transferred to computer for real-time digital human simulation. The data are used for motion type recognition and analysis [5].

Furthermore, video technique is used for observing business life in Japan with great efficiency. It gives help in developing ergonomic methods, statistical methods and technology. Using this approach, the individual user's everyday behaviour is observed as well as employees' collective work, for example during fixing of gas pipelines [6]. As possible outcomes, the satisfaction of consumers and the creation of safe work environment can be increased at the same time.

Video recordings were made during the picking processes in car industry, when several hours of work were recorded [7]. The aim of the recording was to allow employees to mark specific motions in the evaluating program called VIDAR that caused them pain or made them feel uncomfortable. The results showed that after the company received the data they were able to change processes successfully from the ergonomic analysis of complex work.

The video-based method can be applied in all cases, where the goal is to better understand the workflow related processes. The camera-based method has advantages and disadvantages, thus its use needs to be evaluated whether it provides the best solution.

The measurement framework

Goal setting and planning

We designed a process for addressing the questions, such as when we the measured to be done, how and where the measurement can be done for these intervals, what the benefits of the collected information are, and what information we want from the recordings During our research, we based our measurements on a technique utilizing video recordings. Video was recorded of the examined processes and offline analysis was carried out afterwards to extract the required information. The planning phase included decisions on the aforementioned questions, most importantly what to record, how to perform the recording without interfering with the processes and how much captured information is needed to extract certain data.

Implementation

Videos were recorded using commercial HDV camcorders. HDV/AVCHD (1080i50) provides benefit over DV (576i50) due to the higher resolution. Higher image resolution offers more details for analysis. Both systems are capable of interlaced video recording, which offers doubled number of samples at the expense of resolution (row count is halved). This is a practical feature that balances high resolution and larger frame rates. In order to extract accurate time information in both fields of the video may be used.

The recorded videos were transferred to the PC. Tape based camcorders require an IEEE 1394 port to capture the video stream, while recent flash drive based camcorders offer simple file copying mechanism to download video clips. These video files may be transformed and compressed further for archiving purposes or simply fit them to the capacity of the PC used for analysis. We compressed all source videos to a common MPEG4 format to ease software dependencies for decoding.

A custom software application was developed for the evaluation of video recordings. The software enables the user to import a video clip and set up basic parameters of the recording. All these settings may be saved to or load from a project file. Several grid views are provided for the user to define the stages of the analysed process. Each row of the grid defines a process step, each column defines a cycle of the periodic process. The user is able to control freely the video playback (slow down, speed up, rewind, jump to other positions etc.) and at any time simply insert the frame number into the grid or go back to these positions. So as time advances, the user fills up the table from the starting to the finishing stages. After the process is completed all the timing information are available for the process for each cycle and stage. Several tables are available right away to collect information for different (parallel, additional or interrupting) processes. Optional comments or the number of processed products for each cycle can be recorded as annotations.

Data from the grid is transferred to a spreadsheet to analyse the raw data for timing. The pre-defined spreadsheet calculates simple statistical descriptors for the data. In the case redundant observations are available from several cycles, outlier data can be removed and the descriptors can be recalculated. *Fig.2* shows the screenshot of the software.

The spreadsheet calculates a suggested normal cycle time involving extra allowances. The number of cycles and, in case of batch production, the number of product per cycles was also determined. The expected number of products processed in unit time (e.g. hour or shift) is also apparent. This information is essential for production scheduling (see *Fig.3*). The analysing software is available in English and in German.
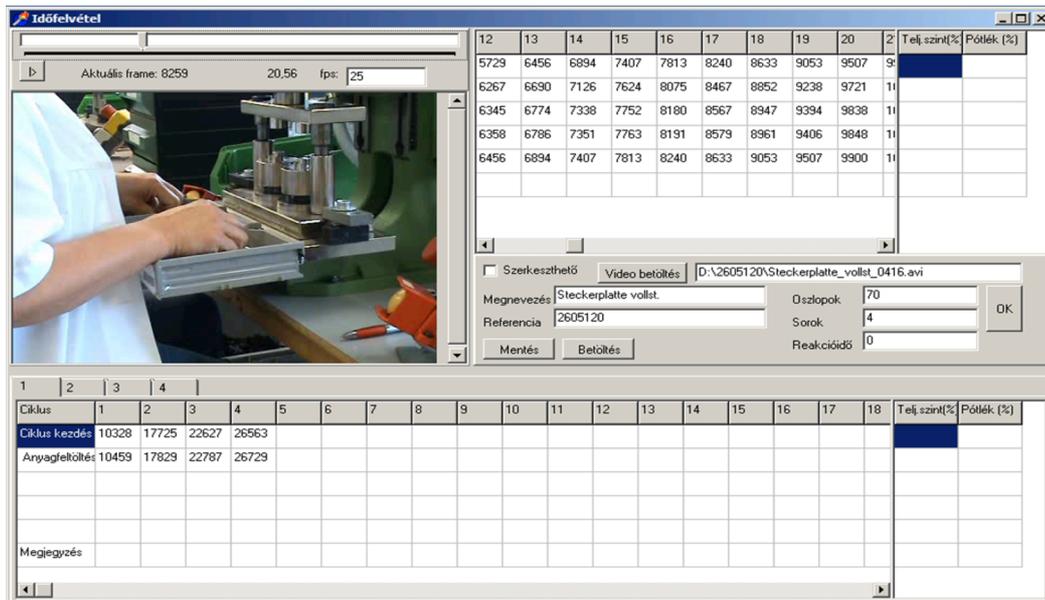
Figure 2: Screenshot of the software

Figure 3: Example for the result of calculation

Figure 4: Dial-type assembly machines at TRANSMODULS Ltd. (graphics were generated by Autodesk Inventor software)

Implementation of analysis

TRANSMODULS Ltd., an engineering and construction company began to fulfil the order of a customer from the automotive industry since April 2012. Mechanical and electrical designs were the first steps of the project. *Fig.4* illustrates the design stage of machines before production. While the mechanical and electrical design tasks were completed, we have gained information about machines. The three individual rotary configuration machines were organized in a production line, thus their operation is not independent from each other. The planned lead-time of the project was 47 weeks that included the mechanical and electrical design, high-value raw materials and components procurement, mechanical and electrical assembly, and installation. Mechanical assembly was followed by programing and the machines were tested during this phase. Some machine movements were already programmed while the operator dependent ones, such as product forwarding were initialized manually. This allowed for adopting to the operator's pace. Later we took the learning curve into account as well.

We used three techniques to obtain time data: (i) Method Time Measurement, MTM, (ii) MAYNARD Operation Sequence Technique, MOST, (iii) video recordings. The two kinds of products to be produced on

Table 1: Summary of data obtained for stations

Method	Cycle time (HD), s	Output, pc/shift
Station 1		
MTM	26.316	830
MOST	24.48	830
Video	97.284	290
Station 2		
MTM	9.2628	910
MOST	9.72	910
Video	16.42	539
Station 3		
MTM	29.226	620
MOST	27.72	620
Video	67.98	415

the production line are heavy duty (HD) and standard (ST). We analysed two areas of the production line: cycle times and capacity. We compared the preliminary traditional calculations (MTM, MOST) and analysis of video recording. We obtained different results *(Table 1)*.

One can observe that cycle times determined by video were much longer than those resulted from MTM and MOST analyses. This can be explained by several reasons, such as the technology is not setup and trained yet, operators are not skilled for the task, supply processes are not ideal, and/or the workplace is not optimized ergonomically.

Table 2: Summary of planning and effective capacity considered ($Q_{planning}$ in piece/hour)

Planning capacity		$Q_{planning}$	Total [a]
Station 1	metal-based	106	4240
Station 1	plastic-based	113	4520
Station 2	metal- and plastic-based	116	4640
Station 3	metal-based	79	3160
Station 3	plastic-based	86	3440
Effective capacity [b]			
Station 1	metal-based	106	3816
Station 1	plastic-based	113	4068
Station 2	metal- and plastic-based	116	4176
Station 3	metal-based	79	2844
Station 3	plastic-based	86	3096

[a] calculated for 8 hours/day and 5 days/week; [b] 90% efficiency is considered in total

Figure 4: A frame from recording for the original layout

Results from MOST and MTM can be seen as target values during operation in ideal case, while video recording shows what production planning and control can expect at start-up operation.

Using the results of time analysis for capacity planning, we are able to forecast the expected value of capacity and capacity utilization independently the source of time data. *Table 2* presents examples for MTM and MOST based analyses of planning capacities and effective capacities. Additional times were taken into consideration resulting in 10% increase in total capacity.

With respect of capacity utilization, the video-based analysis shows that the actual output at least for the first period of operation is lower than the expected output by movement time standard. Using this information, we can calculate expected efficiency, which we extended with 10% rest period time. Later learning for the operator and process improvement will allow for reducing time to complete a workpiece.

Recordings allowed us also to analyse ergonomics of work place *(Fig.4)*. It became obvious at the start of recording that the planned position of the rack is not optimal. Also the upper frame did not allow the operator to use economical movement. These observations suggested constructional changes.

Conclusions

Analysis of workstations in final production phase helps developer, operators, and customers to evaluate a given

Table 3: Summary of capacity utilization and efficiency considered

Capacity utilization		Expected output[a]	Planning capacity[a]	Ratio
St. 1	metal-based	1475	4240	35%
St. 1	plastic-based	1475	4520	33%
St. 2	metal-/plastic-based	8580	4640	184%
St. 3	metal-based	2075	3160	66%
St. 3	plastic-based	2075	3440	60%
Efficiency				
St. 1	metal-based	1475	3816	39%
St. 1	plastic-based	1475	4068	36%
St. 2	metal-/plastic-based	8580	4176	206%
St. 3	metal-based	2075	2844	73%
St. 3	plastic-based	2075	3096	67%

[a] output and capacity are given in pieces/hour

machine in advance before finalizing its design. It allows them to forecast the capacity, capacity utilization, operators' load, and potential ergonomics problems. Tests with operators during the early phase revealed ergonomics problems, such as height, structure and position of hardware elements. Implications of their changes were easily observed repetitively. Irregular activities such as jam cleaning can be analysed also.

REFERENCES

[1] KAMATH M.: Modelling operator/workstation interference in asynchronous automatic assembly systems, Discrete Event Dynamic Systems, *1991*, 5(1), 93–124

[2] ELNEKAVE M., GILAD I.: Rapid video-based analysis system for advanced work measurement, Int. J. Prod. Res., *2006*, 44(2), 271–290

[3] WANG H.C., CHIANG W.C., CHEN S.Y., KE Y.L., CHI C.L., YANG C.W., LIN P.C., KO P.C., WANG Y.C., TSAI T.C., HUANG C.H., HSIUNG K.H., MA M.H., CHEN S.C., CHEN W.J., LIN F.Y.: Video-recording and time-motion analyses of manual *versus* mechanical cardiopulmonary resuscitation during ambulance transport, Resuscitation, *2007*, 74, 453–460

[4] TEIZER J., VELA P.A.: Personnel tracking on construction sites using video cameras, Adv. Engng. Inform., *2009*, 23(4), 452–462

[5] MA L., ZHANG W., FU H., GUO Y., CHABLAT D., BENNIS F., SAWANOI A., FUGIWARA N.: A framework for interactive work design based on motion tracking, simulation, and analysis, Human Factors and Ergonomics in Manufacturing & Service Industries, *2012*, 20, 339–352

[6] MATSUNAMI H.: Service science: applications of observation to real world business, Proc. 4th Int. Conf. Universal Access in Human-Computer Interaction UAHCI'07: Applications and Services, *2007*, pp. 951–960

[7] KADEFORS R., RORSMAN M.: Ergonomic evaluation of complex work: a participative approach employing video-computer interaction, exemplified in a study of order picking, Int. J. Ind. Ergon., *2000*, 25, 435–445

QUASI-POLYNOMIAL CONTROL OF A SYNCHRONOUS GENERATOR

ATTILA MAGYAR[⊠1] AND ATTILA FODOR[1]

[1]Department of Electrical Engineering and Information Systems, University of Pannonia, Egyetem u. 10., H8200 Veszprém,
HUNGARY
⊠E-mail: magyar.attila@virt.uni-pannon.hu

A simple dynamic model of permanent magnet synchronous generator, that is used for electrical energy generation is investigated in this work using a nonlinear technique based on the quasi-polynomial representation of the dynamical model. It is well known that dynamical systems with smooth nonlinearities can be embedded in a quasi-polynomial model. Quasi-polynomial systems are good candidates for a general nonlinear system representation since their global stability analysis is equivalent to the feasibility of a LMI. Moreover, the stabilizing quasi-polynomial state feedback controller design problem is equivalent to the feasibility of a bilinear matrix inequality. The classical stabilizing state feedback problem for quasi-polynomial systems has been extended in this work with the ability of tracking time-dependent reference signals. It is shown, that the stabilizing quasi-polynomial servo controller design is equivalent to a bilinear matrix inequality. The results are applied to the model of a synchronous generator.

Keywords: quasi-polynomial systems, Lotka-Volterra systems, stability analysis, state feedback control, synchronous generator, wind turbine

Introduction

Electrical power systems should operate in an economic way with minimum possible operating cost under normal operating conditions. To ensure this, a preventive controller for power systems has been presented in [1]. It encompasses many types of control actions, including generation rescheduling, load curtailment and network switching reactive compensation.

From the viewpoint of the power grid, the electric power generation can be characterized by the operation of the electrical generators, the subject of our study. These power plants should not only be able to follow the time-varying active power demand of the consumers and the central dispatch center, but also keep the quality indicators (frequency, waveform, total harmonic distortion) of the grid on the expected level. This can be achieved by applying proper control methods based on dynamic models of plant (see e.g. [2, 3]) and the involved generators.

Because of the specialties and great practical importance of the synchronous generators in power plants, their modeling for control purposes is also well investigated in the literature. Besides of the basic textbooks (see e.g. [4]) that describe the modeling, specialized papers are also available that use the developed models for the design of various controllers [5].

A wind turbine driving permanent magnet synchronous generator is proposed in [6] with current controlled voltage source inverter, which is the best choice when the output power is small. The current control of the voltage source inverter has bidirectional active and reactive power control ability, avoiding the intricacy of the controller designing.

The class of quasi-polynomial (QP) systems plays an important role in the theory of nonlinear dynamical systems because nonlinear systems with smooth nonlinearities can be transformed into a QP form [7]. This means, that any applicable method for QP systems can be regarded as a general technique for nonlinear systems.

Previous work in the field of QP systems include [8], which proves that the global stability analysis of QP systems is equivalent to the feasibility of a linear matrix inequality (LMI). It has been shown in [9] that the globally stabilizing state feedback design for QP systems is equivalent to a bilinear matrix inequality (BMI). Although the solution of a BMI is an NP-hard problem, an iterative LMI algorithm could be used. A summary of linear and bilinear matrix inequalities and the available software tools for solving them can be found in [10]. Another control synthesis algorithm for polynomial systems is presented in [11].

The goal of this paper is to formulate the servo controller design problem for QP systems based on the results presented in [9] and to design a servo controller for a synchronous generator model using the QP controller synthesis methodology that keeps the active power at the desired level.

Basic notions

In what follows, the basic modeling assumptions and definitions to be used in the sequel are summarized briefly.

Nonlinear model of a synchronous generator

The modeling procedure of the synchronous generator is mainly based on [12] and [13], therefore, only the resulting model is presented here.

The model is based on the following simplification assumptions:

- a symmetrical tri-phase stator winding system is assumed,
- one field winding is considered to be in the machine,
- all of the windings are magnetically coupled,
- the flux linkage of the winding is a function of rotor position,
- the copper loss and the slots in the machine can be neglected,
- the spatial distribution of the stator fluxes and apertures wave are considered to be sinusoidal, and
- stator and rotor permeability are assumed to be infinite.

It is also assumed that all the losses due to wiring, saturation, and slots can be neglected.

The four windings (three stators and one rotor) are magnetically coupled. Since the magnetic coupling between the windings is a function of the rotor position, the flux linkage of the windings is also a function of the rotor position. The actual terminal voltage v of the windings can be written in the form

$$v = \pm \sum_{j=1}^{J} (r_j\, i_j) \pm \sum_{j=1}^{J} \left(\frac{\mathrm{d}\varphi_j}{\mathrm{d}t} \right), \qquad (1)$$

where i_j are the currents, r_j are the winding resistances, and φ_j are the flux linkages. The positive directions of the stator currents point out of the synchronous generator terminals.

Thereafter, the two stator electromagnetic fields, both traveling at rotor speed, can be identified by decomposing each stator phase current under steady state into two components, one in phase with the electromagnetic field and another phase shifted by 90°. With the above, one can construct an air-gap field with its maximal aligned to the rotor poles (d axis), while the other is aligned to the q axis (between poles). This method is called the Park's transformation.

As a result of the derivation in [12] the vector voltage equation is as follows

$$\boldsymbol{v}_{dFq} = -\boldsymbol{R}\,\boldsymbol{i}_{dFq} - \boldsymbol{L}\,\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{i}_{dFq}, \qquad (2)$$

with

$$\begin{aligned} \boldsymbol{v}_{dFq} &= \begin{bmatrix} v_d & -v_F & v_q \end{bmatrix}^T \\ \boldsymbol{i}_{dFq} &= \begin{bmatrix} i_d & i_F & i_q \end{bmatrix}^T, \end{aligned} \qquad (3)$$
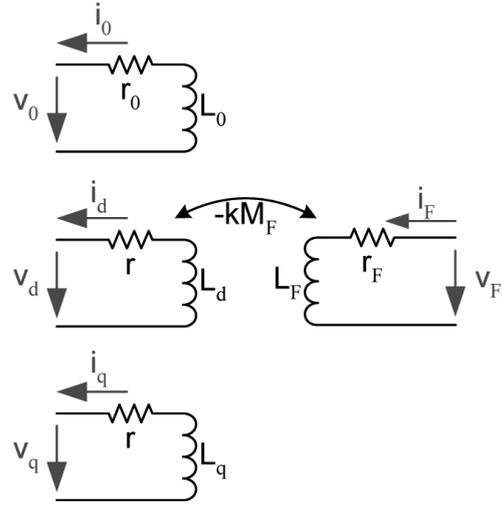
Figure 1: The equivalent circuit of the synchronous generator.

where v_d, v_q and i_d, i_q are the direct and the quadrature components of the stator voltage and current of the synchronous generator, while v_F and i_F are the exciter voltage and current of the synchronous generator. Furthermore, $\boldsymbol{R}$ and $\boldsymbol{L}$ are the following matrices (see *Fig.*1)

$$\begin{aligned} \boldsymbol{R}_{RS\omega} &= \begin{bmatrix} r & 0 & \omega\, L_q \\ 0 & r_F & 0 \\ -\omega\, L_d & -\omega\, k\, M_F & r \end{bmatrix} \\ \boldsymbol{L} &= \begin{bmatrix} L_d & k\, M_F & 0 \\ k\, M_F & L_F & 0 \\ 0 & 0 & L_q \end{bmatrix}, \end{aligned} \qquad (4)$$

where r is the stator resistance, r_F is the exciter resistance of the Synchronous Generator, L_d, and L_q are the direct and the quadrature part of the stator and rotor inductance, ω is the angular velocity, and M_F is linkage inductances. The state-space model for the currents is obtained by expressing $\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{i}_{dFq}$ from *Eq. (2)*, i.e.

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{i}_{dFq} = -\boldsymbol{L}^{-1}\,\boldsymbol{R}_{RS\omega}\,\boldsymbol{i}_{dFq} - \boldsymbol{L}^{-1}\,\boldsymbol{v}_{dFq}. \qquad (5)$$

The purely electrical model *Eq. (5)* has to be extended with the equation of rotational motion (*Eq. (6)*) that gives the mechanical sub-dynamics, that is

$$\begin{aligned} \frac{\mathrm{d}\omega}{\mathrm{d}t} = &-\frac{L_d\, i_q}{3\tau_j}i_d + \frac{-k\, M_F\, i_q}{3\tau_j}i_F + \\ &+ \frac{L_q\, i_d}{3\tau_j}i_q + \frac{-D}{\tau_j}\omega + \frac{T_{\mathrm{mech}}}{\tau_j}. \end{aligned} \qquad (6)$$

$$\begin{bmatrix} \dot{i}_d \\ \dot{i}_F \\ \dot{i}_q \\ \dot{\omega} \end{bmatrix} =$$

$$\begin{bmatrix} -\frac{L_F}{H} & \frac{k M_F}{H} & 0 & 0 \\ \frac{k M_F}{H} & -\frac{L_d}{H} & 0 & 0 \\ 0 & 0 & \frac{1}{L_q} & 0 \\ 0 & 0 & 0 & \frac{1}{\tau_j} \end{bmatrix} \begin{bmatrix} -v_d \\ v_F \\ -v_q \\ T_{mech} \end{bmatrix} +$$

$$\begin{bmatrix} -\frac{rL_F}{H} & \frac{kM_Fr_F}{H} & -\frac{\omega L_F L_q}{H} & 0 \\ \frac{rkM_F}{H} & -\frac{r_F L_d}{H} & \frac{\omega kM_F L_q}{H} & 0 \\ -\frac{\omega L_d}{L_q} & \frac{\omega kM_F}{L_q} & \frac{r}{L_q} & 0 \\ -\frac{L_d\,i_q}{3\tau_j} & \frac{-k\,M_F\,i_q}{3\tau_j} & \frac{L_q\,i_d}{3\tau_j} & \frac{-D}{\tau_j} \end{bmatrix} \begin{bmatrix} i_d \\ i_F \\ i_q \\ \omega \end{bmatrix} \quad (7)$$

where

$$H = k^2 M_F^2 - L_d L_F$$

$$\begin{bmatrix} \dot{i}_d \\ \dot{i}_F \\ \dot{i}_q \\ \dot{\omega} \end{bmatrix} =$$

$$\begin{bmatrix} -1.7100 & 0.5893 & 0 & 0 \\ 0.5893 & -6.6918 & 0 & 0 \\ 0 & 0 & -1.7090 & 0 \\ 0 & 0 & 0 & 0.0006 \end{bmatrix} \begin{bmatrix} -v_d \\ v_F \\ -v_q \\ T_{mech} \end{bmatrix} +$$

$$\begin{bmatrix} -0.0019 & 0.0004 & -3.4883\,\omega & 0 \\ 0.0006 & -0.0042 & 1.2022\,\omega & 0 \\ 3.5888\,\omega & 2.6489\,\omega & -0.0019 & 0 \\ -0.0004\,i_q & -0.0003\,i_q & 0.0004\,i_d & -0.0011 \end{bmatrix} \begin{bmatrix} i_d \\ i_F \\ i_q \\ \omega \end{bmatrix}$$

(8)

Altogether, the state equations *Eqs.* (5) and (6) have four state variables: i_d, i_F, i_q, and ω.

The manipulated input vector of the generator is $\boldsymbol{u} = \begin{bmatrix} v_F & T_{mech} \end{bmatrix}^T$, while the disturbance input vector is $\boldsymbol{d} = \begin{bmatrix} v_d & d_q \end{bmatrix}^T$. Realize that the state equations *Eqs.* (5) and (6) are bilinear in the state variables because matrix $\boldsymbol{R_{RS\omega}}$ depends linearly on ω. The obtained model is nonlinear and it has four state variables: i_d, i_F, i_q and ω [14].

Output equations of the model: The output active power equation can be written in the following form:

$$p_{out} = v_d i_d + v_q i_q \qquad (9)$$

and the reactive power is

$$q_{out} = v_d i_q - v_q i_d. \qquad (10)$$

Eqs. (9) and (10) are the output equations of the generator's state space model. It is important to note, that these equations are bi-linear in the state and input variables. Note, that although only the active power is to be controlled in this case, as a possible future extension of the work, reactive power can also be controlled in order to follow an external reference signal.

Quasi-polynomial representation of nonlinear systems

Let us denote the element of an arbitrary matrix $\boldsymbol{W}$ with row index i and column index j by W_{ij}. Quasi-polynomial models are systems of ordinary differential equations (ODE) of the following form

$$\dot{y}_i = y_i \left(L_i + \sum_{j=1}^{m} A_{ij} \prod_{k=1}^{n} y_k^{B_{jk}} \right), \quad i = 1, \ldots, n.$$

(11)

where $y \in \text{int}(\mathbb{R}_+^n)$, $A \in \mathbb{R}^{n \times m}$, $B \in \mathbb{R}^{m \times n}$, $L_i \in \mathbb{R}$, $i = 1, \ldots, n$. Furthermore, $\boldsymbol{L} = [L_1 \ \ldots \ L_n]^T$. Let us denote the equilibrium point of interest of *Eq. (11)* as $\boldsymbol{y^*} = [y_1^* \ y_2^* \ \ldots \ y_n^*]^T$. Without the loss of generality we can assume that $\text{rank}(\boldsymbol{B}) = n$ and $m \geq n$ (see [15]).

Lotka-Volterra models

The above family of models is split into classes of equivalence [16] according to the values of the products $\boldsymbol{M} = \boldsymbol{B}\,\boldsymbol{A}$ and $\boldsymbol{N} = \boldsymbol{B}\,\boldsymbol{L}$. The LOTKA-VOLTERRA form gives the representative elements of these classes of equivalence. If $\text{rank}(\boldsymbol{B}) = n$, then the set of ODEs in *Eq. (11)* can be embedded into the following *m*-dimensional set of equations, the so called LOTKA-VOLTERRA model:

$$\dot{z}_j = z_j \left(N_j + \sum_{i=1}^{m} M_{ji} z_i \right), \quad j = 1, \ldots, m \quad (12)$$

where

$$\boldsymbol{M} = \boldsymbol{B}\,\boldsymbol{A}, \quad \boldsymbol{N} = \boldsymbol{B}\,\boldsymbol{L},$$

and each z_j represents a so called quasi-monomial:

$$z_j = \prod_{k=1}^{n} y_k^{B_{jk}}, \qquad j = 1, \ldots, m. \qquad (13)$$

Stability analysis using linear matrix inequalities

Henceforth it is assumed that $\boldsymbol{y}^*$ is a positive equilibrium point, i.e. $\boldsymbol{y}^* \in \text{int}(\mathbb{R}_+^n)$ in the QP case and similarly $\boldsymbol{z}^* \in \text{int}(\mathbb{R}_+^m)$ is a positive equilibrium point in the LOTKA-VOLTERRA case. For LV systems there is a well known candidate LYAPUNOV function family ([8, 17]), which is in the form:

$$V(\boldsymbol{z}) = \sum_{i=1}^{m} c_i \left(z_i - z_i^* - z_i^* \ln \frac{z_i}{z_i^*} \right), \qquad (14)$$

$$c_i > 0, \quad i = 1 \ldots m,$$

where $\boldsymbol{z}^* = \begin{bmatrix} z_1^* & \ldots & z_m^* \end{bmatrix}^T$ is the equilibrium point corresponding to the equilibrium $\boldsymbol{y}^*$ of the original QP system (*Eq. (11)*). The time derivative of the LYAPUNOV function *Eq. (14)* is:

$$\dot{V}(\boldsymbol{z}) = \frac{1}{2}(\boldsymbol{z} - \boldsymbol{z}^*)(\boldsymbol{C}\,\boldsymbol{M} + \boldsymbol{M}^T\boldsymbol{C})(\boldsymbol{z} - \boldsymbol{z}^*), \quad (15)$$

where $\boldsymbol{C} = \text{diag}(c_1, \ldots, c_m)$ and $\boldsymbol{M}$ is the invariant characterizing the LOTKA-VOLTERRA form (*Eq. (12)*). Therefore, the non-increasing nature of the LYAPUNOV function is equivalent to a feasibility problem over the following set of LMI constraints (see [18] or [19]):

$$\begin{aligned} \boldsymbol{C}\,\boldsymbol{M} + \boldsymbol{M}^T\boldsymbol{C} &\leq \boldsymbol{0} \\ \boldsymbol{C} &> \boldsymbol{0} \end{aligned} \qquad (16)$$

where the unknown matrix is $\boldsymbol{C}$, which is diagonal and contains the coefficients of *Eq. (14)*.

The derivation of global stability analysis for nonautonomous QP systems from the autonomous case is straightforward. The LYAPUNOV function (*Eq. (15)*) also depends on the equilibrium value of the input (u^*) and has the form

$$\dot{V}(z) = \frac{1}{2}(z - z^*)(C\tilde{M} + \tilde{M}^T C)(z - z^*), \quad (17)$$

where $\tilde{M}$ depends on the coefficient matrices of the input-affine LOTKA-VOLTERRA model (*Eq. (21)*):

$$\tilde{M} = M_0 + \sum_{l=1}^{p} M_l u_l^*.$$

The corresponding LMI feasibility problem to be solved in order to check global asymptotic stability is

$$\begin{array}{rcl} C\tilde{M} + \tilde{M}^T C & \leq & 0 \\ C & > & 0. \end{array} \quad (18)$$

Input-affine QP system models

An input-affine nonlinear system model with state vector y, input vector u and output vector η

$$\begin{array}{rcl} \dot{y} & = & f(y) + \sum_{i=1}^{p} g_i(y) u_i \\ \eta & = & h(y) \end{array} \quad (19)$$

is in QP-form if all of the functions f, g_i and h are in QP-form. Then the general form of the state equation of an input-affine QP system model with p-inputs is:

$$\dot{y}_i = y_i \left(L_{0_i} + \sum_{j=1}^{m} A_{0_{ij}} \prod_{k=1}^{n} y_k^{B_{jk}} \right) + \quad (20)$$

$$+ \sum_{l=1}^{p} y_i \left(L_{l_i} + \sum_{j=1}^{m} A_{l_{ij}} \prod_{k=1}^{n} y_k^{B_{jk}} \right) u_l$$

where

$$i = 1, \ldots, n, \quad A_0, A_l \in \mathbb{R}^{n \times m}, \quad B \in \mathbb{R}^{m \times n},$$
$$L_0, L_l \in \mathbb{R}^n, \quad l = 1, \ldots, p.$$

The corresponding input-affine LOTKA-VOLTERRA model is in the form

$$\dot{z}_j = z_j \left(N_{0_j} + \sum_{k=1}^{m} M_{0_{jk}} z_k \right) + \quad (21)$$

$$+ \sum_{l=1}^{p} z_j \left(N_{l_j} + \sum_{k=1}^{m} M_{l_{jk}} z_k \right) u_l$$

where

$$j = 1, \ldots, m, \quad M_0, M_l \in \mathbb{R}^{m \times m},$$
$$N_0, N_l \in \mathbb{R}^m, \quad l = 1, \ldots, p,$$

and the parameters can be obtained from the input-affine QP system's ones in the following way

$$\begin{array}{rcl} M_0 & = & B A_0 \\ N_0 & = & B L_0 \\ M_l & = & B A_l \\ N_l & = & B L_l \end{array} \quad l = 1, \ldots, p. \quad (22)$$

The controller design problem for QP systems

Globally stabilizing QP state feedback design problem for QP systems can be formulated as follows (for a more detailed desription, see [9]). Consider arbitrary QP inputs in the form:

$$u_l = \sum_{i=1}^{r} k_{il} \hat{q}_i, \quad l = 1 \ldots, p, \quad (23)$$

where $\hat{q}_i = \hat{q}_i(y_1, \ldots, y_n)$, $i = 1, \ldots, r$ are arbitrary quasi-monomial functions of the state variables of *Eq. (20)* and k_{il} is the constant gain of the quasi-monomial function $\hat{q}_i$ in the l-th input u_l. The closed loop system will also be a QP system with matrices

$$\hat{A} = A_0 + \sum_{l=1}^{p} \sum_{i=1}^{r} k_{il} A_{il}, \quad \hat{B},$$

$$\hat{L} = L_0 + \sum_{l=1}^{p} \sum_{i=1}^{r} k_{il} L_{il}. \quad (24)$$

Note that the number of quasi-monomials in the closed-loop system (i.e. the dimension of the matrices) together with the matrix $\hat{B}$ may significantly change depending on the choice of the feedback structure, i.e. on the quasi-monomial functions $\hat{q}_i$.

Furthermore, the closed loop LV coefficient matrix $\hat{M}$ can also be expressed in the form

$$\hat{M} = \hat{B}\hat{A} = M_0 + \sum_{l=1}^{p} \sum_{i=1}^{r} k_{il} M_{il}. \quad (25)$$

Then the global stability analysis of the closed loop system with unknown feedback gains k_{il} leads to the following BMI

$$\hat{M}^T C + C\hat{M} = M_0^T C + CM_0 + \quad (26)$$

$$\sum_{l=1}^{p} \sum_{i=1}^{r} k_{il} \left(M_{il}^T C + CM_{il} \right) \leq 0.$$

The variables of the BMI are the $p \times r$ k_{il} feedback gain parameters and the c_j, $j = 1, \ldots, m$ parameters of the LYAPUNOV function. If the BMI above is feasible, there exists a globally stabilizing feedback with the selected structure.

Note that (marginal) infeasibility of the BMI (*Eq. (26)*) means only that the closed loop system is not proven to be globally asymptotically stable. However, the solution k_{il} may still guarantee local stability, which is enough in several cases.

Controller design using bilinear matrix inequalities

A BMI is a diagonal block composed of q matrix inequalities of the following form

$$G_0^i + \sum_{k=1}^{p} x_k G_k^i + \sum_{k=1}^{p} \sum_{j=1}^{p} x_k x_j K_{kj}^i \leq 0, \quad i = 1, \ldots, q \quad (27)$$

where $x \in \mathbb{R}^p$ is the decision variable to be determined and G_k^i, $k = 0, \ldots, p$, $i = 1, \ldots, q$ and K_{kj}^i, $k, j = 1, \ldots, p$, $i = 1, \ldots, q$ are symmetric, quadratic matrices.

The main properties of BMIs are that they are non-convex in x (which makes their solution numerically much more complicated than that of linear matrix inequalities), and their solution is NP-hard [10], so the size of the tractable problems is limited. However, there exist practically applicable and effective algorithms for BMI solution [20], [21], or [22]. In Matlab environment the TomLab/PENBMI solver [23] can be used effectively to solve BMIs.

Quasi-polynomial servo control

As in the linear case, the problem statement of the servo, or reference tracking control is as follows. Consider a nonlinear system in the form of *Eq. (20)* and an external reference signal r which is to be followed by the system output η.

It is possible to define the tracking error signal (*Eq. (28)*) whose time derivative gives the tracking error dynamics which should be stabilized together with the system (*Eq. (20)*).

$$z(t) = \int_{t_0}^{t} r(\tau) - \eta(\tau)\mathrm{d}\tau. \tag{28}$$

It is easy to see that the differential equation of the tracking error has the form (*Eq. (29)*).

$$\dot{z}(t) = r(t) - \eta(t), \tag{29}$$

$$u(t) = -K_R \int_0^t r(\tau) - \eta(\tau)d\tau - \boldsymbol{K}\,\boldsymbol{y}(t).$$

If the output equation is also in QP form, then the extended closed loop QP system can be written up in LOTKA-VOLTERRA form similar to (*Eq. (25)*) and the BMI for the globally stabilizing controller design can be formulated.

Quasi-polynomial control of the synchronous generator

Quasi-polynomial form of the synchronous generator

The bilinear nature of the state equations (*Eq. (8)*) enables us to directly apply the QP mechanism without QP embedding [16]. The system has the following set of quasi-monomials:

$$\left\{ \frac{1}{i_d}, \frac{i_F}{i_d}, \frac{i_q\,\omega}{i_d}, \frac{1}{i_F}, \frac{i_d}{i_F}, \frac{i_q\,\omega}{i_F}, \frac{1}{i_q}, \frac{i_d\,\omega}{i_q}, \frac{i_F\,\omega}{i_q}, \frac{1}{\omega}, \frac{i_F\,i_q}{\omega} \right\}$$

The QP coefficient matrices of the input-affine system (*Eq. (20)*) are

$$\boldsymbol{A_0} = \begin{bmatrix} -2.323 & 0 & 0 & 0 \\ 0.0004 & 0 & 0 & 0 \\ 0.5893 & 0 & 0 & 0 \\ -3.4883 & 0 & 0 & 0 \\ 0 & 0.8 & 0 & 0 \\ 0 & 0.0006 & 0 & 0 \\ 0 & -6.6918 & 0 & 0 \\ 0 & 1.2022 & 0 & 0 \\ 0 & 0 & -0.7861 & 0 \\ 0 & 0 & 3.5888 & 0 \\ 0 & 0 & 2.6489 & 0 \end{bmatrix}^T$$

$$\boldsymbol{A_1} = \begin{bmatrix} 0.5893 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -6.6918 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}^T$$

$$\boldsymbol{A_2} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.0006 \end{bmatrix}^T$$

Feedback structure

There is a degree of freedom in the selection of the stabilizing feedback structure. A wise choice of the feedback structure does not increase the number of monomials of the closed loop system. This way the size of the BMI to be solved remains tractable. In our case, a linear full state feedback is applied, i.e. the feedback law is in the form

$$\begin{aligned} v_F \;=\; & k_{r1} \int_0^t r(\tau) - p_{out}(\tau)d\tau + \\ & + k_1\,i_d + k_2\,i_F + k_3\,i_q + k_4\,\omega \\ T_{mech} \;=\; & k_{r2} \int_0^t r(\tau) - p_{out}(\tau)d\tau + \\ & + k_5\,i_d + k_6\,i_F + k_7\,i_q + k_8\,\omega. \end{aligned} \tag{30}$$

Controller design and verification via simulation

Using the feedback law (*Eq. (30)*), the closed loop system is also in QP form with 18 quasi-monomials:

$$\left\{ \frac{1}{i_d}, \frac{i_F}{i_d}, \frac{i_q}{i_d}, \frac{\omega}{i_d}, \frac{i_q\omega}{i_d}, \frac{1}{i_F}, \frac{i_d}{i_F}, \frac{i_q}{i_F}, \frac{\omega}{i_F}, \frac{i_q\omega}{i_F}, \frac{1}{i_q}, \frac{i_d\omega}{i_q}, \right.$$

Figure 2: The controlled output p_{out} of the system (solid line) together with the reference input r (dashed line).

$$\left. \frac{i_F\omega}{i_q}, \frac{i_d}{\omega}, \frac{i_F}{\omega}, \frac{i_q}{\omega}, \frac{i_di_q}{\omega}, \frac{i_Fi_q}{\omega} \right\}.$$

Due to the lack of space, the 19×19 (the extra dimension comes from the tracking error dynamics) matrices of the BMI and the closed loop LOTKA-VOLTERRA system are not listed here.

The BMI (*Eq. (26)*) for the globally stabilizing feedback design case suffers from rank deficiency and the available BMI solvers stop with marginal infeasibility, which means that global stability cannot be proven using the feedback law (*Eq. (30)*). However, simulation results indicate, that the system is locally stable with the controller gain parameters yielded by the globally stabilizing BMI (*Eq. (26)*) formulated for the closed loop generator model extended with the tracking error dynamics.

$$
\begin{aligned}
k_{r1} &= 1.0000 & k_{r2} &= -0.0002 \\
k_1 &= -0.2932 & k_2 &= -1.6880 \\
k_3 &= -1.9609 & k_4 &= 0.2259 \\
k_5 &= 0.0001 & k_6 &= 0.0000 \\
k_7 &= -0.0001 & k_8 &= 0.1450
\end{aligned}
$$

The simulated behavior of the controlled generator can be seen in *Fig.* 2 where the controlled output of the system (i.e. p_{out}) is shown together with the reference input (r). Both of them are dimensionless. The tracking properties of the system are acceptable, moreover, its disturbance rejection is also good (at time 40, a step-like change has been applied to the disturbance input v_d).

The simulation has been performed in Matlab/Simulink environment [24].

Conclusions

A novel servo control design technique based on the QP representation has been formulated in this work. As an example, an active power tracking controller has been designed for a synchronous generator model. The tracking properties of the closed loop system are satisfactory.

Further work includes the extension of the method for vector reference signals and use the degree of freedom lying in the BMI problem for formulating a robust/optimal controller design problem. Another direction of future research is to apply graph theoretical techniques for controller structure selection that applies the underlying connections between QP systems and chemical reaction networks [25].

REFERENCES

[1] VERMA, K., NIAZI, K. A coherency based generator rescheduling for preventive control of transient stability in power systems, International Journal of Electrical Power & Energy Systems, *2013*, 45, 10–18

[2] BANAVAR, R., DESHPANDE, U. Robust controller design for a nuclear power plant using H-infinity optimization, in Proc. of the 35th Conference on Decision and Control, Kobe, Japan

[3] NA, M., JUNG, D., SHIN, S., JANG, J., LEE, K., LEE, Y.: A model predictive controller for loadfollowing operation of PWR reactors, IEEE Transactions on Nuclear Science, *2005*, 52,1009–1020

[4] ANDERSON, P., FOUAD, A. Power-Systems-Control and Stability (The IOWA State University Press, Ames Iowa, USA), *1977*

[5] DE MELLO, F., CONCORDIA, C. Concepts of synchronous machine stability as affected by excitation control, IEEE Transactions on Power Apparatus and Systems, *1969*, PAS-88(4), 316–329

[6] MUYEEN, S., AL-DURRA, A., TAMURA, J. Variable speed wind turbine generator system with current controlled voltage source inverter, Energy Convers Manage, *2011*, 52

[7] HERNÁNDEZ-BERMEJO, B., FAIRÉN, V. Nonpolynomial vector fields under the Lotka-Volterra normal form, Physics Letters A, *1995*, 206, 31–37

[8] FIGUEIREDO, A., GLERIA, I. M., ROCHA, T. M. Boundedness of solutions and Lyapunov functions in quasi-polynomial systems, Physics Letters A, *2000*, 268, 335–341

[9] MAGYAR, A., SZEDERKÉNYI, G., HANGOS, K. M. Globally stabilizing feedback control of process systems in generalized Lotka-Volterra form, Journal of Process Control, *2008*, 18(1), 80–91

[10] VANANTWERP, J., BRAATZ, R. A tutorial on linear and bilinear matrix inequalities, Journal of Process Control, *2000*, 10, 363–385

[11] TONG, C.-F. , ZHANG, H., SUN, Y.-X. Controller Design for Polynomial Nonlinear Systems

with Affine Uncertain Parameters, Acta Automatica Sinica, *2007*, 33(12), 1321–1325

[12] FODOR, A., MAGYAR, A., HANGOS, K. M. Control-Oriented Modeling of the Energy-Production of a Synchronous Generator in a Nuclear Power Plant, Energy, *2012*, 39, 135–145

[13] FODOR, A., MAGYAR, A., HANGOS, K. M. Parameter Sensitivity Analysis of a Synchronous Generator, Hungarian Journal of Industrial Chemistry, *2010*, 38(1), 21–26

[14] FODOR, A., MAGYAR, A., HANGOS, K. Dynamic modeling and model analysis of a large industrial synchronous generator, in Applied Electronics, Pilsen, Czech Republic, 91–96

[15] HERNÁNDEZ-BERMEJO, B., FAIRÉN, V., BRENIG, L. Algebraic recasting of nonlinear ODEs into universal formats, J. Phys. A, Math. Gen., *1998*, 31, 2415–2430

[16] HERNÁNDEZ-BERMEJO, B., FAIRÉN, V. Lotka-Volterra representation of general nonlinear systems, Math. Biosci., *1997*, 140, 1–32

[17] HERNÁNDEZ-BERMEJO, B. Stability conditions and Lyapunov functions for quasi-polynomial systems, Applied Mathematics Letters, *2002*, 15, 25–28

[18] BOYD, S., GHAOUI, L. E. , FERON, E., BALAKRISHNAN, V. Linear Matrix Inequalities in System and Control Theory (SIAM, Philadelphia) *1994*

[19] SCHERER, C., WEILAND, C. Linear Matrix Inequalities in Control (DISC, http://www.er.ele.tue.nl/sweiland/lmi.pdf), *2000*

[20] KOCVARA, M., STINGL, M. A code for convex nonlinear and semidefinite programming, Optimization Methods and Software, *2003*, 8, 317–333

[21] TUAN, H., APKARIAN, P., NAKASHIMA, Y. A new Lagrangian dual global optimization algorithm for solving bilinear matrix inequalities, International Journal of Robust and Nonlinear Control, *2000*, 10, 561–578

[22] CAO, Y.-Y., LAM, J., SUN, Y.-X. Static Output Feedback Stabilization: An ILMI Approach, Automatica, *1998*, 12, 1641–1645

[23] KOCVARA, M., STINGL, M. TOMLAB/PENBMI solver (Matlab Toolbox) *2005*, PENOPT Gbr.

[24] I. THE MATHWORKS: MATLAB environment *2007*) copyright 1984-2007 The MathWorks, Inc. http://www.mathworks.com

[25] HANGOS, K. M., SZEDERKÉNYI, G. The underlying linear dynamics of some positive polynomial systems, Physics Letters A, *2012*, 376(45), 3129–3134

Advertise upcoming meetings,

conferences, and workshops;

make public announcements;

introduce your research laboratory;

a new product or a service

in the

Hungarian Journal of Industry and Chemistry

Please contact us if interested!

A STRUCTURED MODEL-BASED DIAGNOSIS METHOD FOR DISCRETE DYNAMIC PROCESSES USING EVENT SEQUENCES

Attila Tóth [✉][1], Katalin M. Hangos[1,2], and Ágnes Werner-Stark[1]

[1]Department of Electrical Engineering and Information Systems, University of Pannonia, Veszprém H-8200, HUNGARY
[2]Computer and Automation Research Institute, Budapest, HUNGARY
✉ E-mail: atezs82@gmail.com

A novel model-based fault detection and diagnosis method is proposed that is based on following event sequences measured in a discrete dynamic process. The model of the nominal and faulty operation modes is given in the form of event sequences, that are decomposed according to the components and sub-components present in the process system. The faulty event sequences are defined using extended procedure HAZID tables. A diagnostic algorithm is also presented that uses a component-wise decomposed form of the event sequences. The operation of the algorithm is illustrated on a simple example of a process system consisting of three similar tanks.

Keywords: Process monitoring, diagnosis, discrete event systems, qualitative models

Introduction

Fault prevention and mitigation in the field of process system management is a task of crucial importance in avoiding serious accidents. Thus, numerous hazard identification (HAZID) techniques have been developed in the past decades to ensure the safe operation of process systems and to relieve effects of faults (see [1] for a broad presentation of the field). Among these techniques, the most important methodologies involve the function-driven HAZOP (HAZard and OPerability, see [2]) analysis and the component-driven FMEA (Fault Mode and Effects Analysis). There have been results in the past decade for automating the creation of HAZOP analysis ([3] with a concrete application described in [4]). Blending the component-driven and function-driven analyses also resulted in a novel hazard identification approach described in [5].

Although the information collected in the HAZOP and FMEA studies serve the purpose of hazard identification, these studies can be the basis of diagnostic procedures, too. A model-based diagnostic method based on HAZOP and FMEA information is reported in [6].

It is important to note that the above techniques concentrate on the static case when the deviation from a normal steady-state behaviour is of importance. Therefore, the transient case when the plant is controlled by an operational procedure is not addressed in these results. A recent study [7] tries to deal with the diagnostic task by using a specially constructed P-HAZID analysis and a diagnostic algorithm. In this paper, this diagnostic idea is extended to be able to handle more complex diagnostic tasks - by taking advantage of a possible decomposition of typical process systems along their similar components.

Basic notions

In a complex system the full dynamic model that describes its behaviour under normal and faulty operation models is rarely available, therefore one should base the diagnosis on qualitative information both in terms of the dynamic models and in the measured data. Here we briefly summarize the basic notions for qualitative model based diagnosis.

Qualitative range spaces

Current values of continuous measurable outputs in process systems can be described using a properly selected qualitative range space. For example, to describe the value of a level sensor in a tank, the following range space can be used:

$$Q_e = \{\mathbf{e}-, 0, L, N, H, \mathbf{e}+\}. \tag{1}$$

Here, 0 means an empty tank, L, N and H means low, normal and high fluid level, respectively, while $\mathbf{e}-$ and $\mathbf{e}+$ refer to unmeasurably low and high fluid levels (this might mean a failure in the level sensor itself). This range space will be used to describe system outputs during operation.

Input-output event sequences

Operational procedures in process systems are detailed list of instructions for the plant operator personnel to

perform certain operations on the plant. Procedures can be formally described using finite input-output event sequences where a single event describes a change in either the inputs or the outputs of the system at a specific time instant. Therefore the syntax of a single input-output event (at time instant t) is the following

$$\text{event}_t = (t; \text{input values}; \text{output values}).$$

The input in an event always refers to a state of an actuator component in the process system (e.g., in the case of a valve it can be **open** (op) or **closed** (cl)). On the other hand, the output in an event refers to a value of an output of the process system in the qualitative range space using the qualitative set defined in *Eq.* (1). Sequences formed from these events are called traces and defined as

$$T(t_1, t_n) = \text{event}_{t_1}, ..., \text{event}_{t_n}.$$

Separate events in a trace contain the same inputs and outputs. Note that the discrete event time instances t_i are abbreviated by their indexes i in the description, i.e. 2 stands for t_2.

Examples of single events for a two-input single output case include

$$(1; \text{cl}, \text{op}; \text{N}) \quad , \quad (2; \text{cl}, \text{op}; \text{L})$$

where at time instances t_1 and t_2 the two valve inputs are held closed and open, respectively, while the level decreases form its normal value to the low level. The trace formed form the above two consecutive events is written as

$$T(1, 2) = ((1; \text{cl}, \text{op}; \text{N}), (2; \text{cl}, \text{op}; \text{L})).$$

For every operational procedure there exists a trace (called the nominal trace) which describes its behaviour under fault-free conditions. The diagnostic method compares this trace to other traces which may have been executed under faulty conditions (called characteristic traces), and the differences (called deviations) are later used to find possible malfunctions of components in the system.

Deviations

Nominal and characteristic traces can be compared by comparing their corresponding event fragments. The difference between two corresponding event fragments is described by a deviation. Deviations are formed from a deviation guideword and the nominal event from which the corresponding characteristic trace event is deviating from. The following deviation types are used during diagnosis:

- **never-happened**: When the particular event never happened in the characteristic trace.
- **later**: When the event happened in the characteristic trace, but at a later time instant.

- **earlier**: When the event happened in the characteristic trace, but at an earlier time instant.
- **greater**: When a particular output's qualitative value was higher in the characteristic event than that of the nominal trace.
- **smaller**: When a particular output's qualitative value was lower in the characteristic event than that of the nominal trace.

For the detailed description of the **greater** and **smaller** qualitative relations, please refer to [7].

Procedure HAZID

As a combination and extension of the widely used FMEA and HAZOP analyses (for details, refer to [7] or [1], and in particular to [5]), the procedure HAZID (abbreviated as P-HAZID) analysis can be used for fault diagnosis during operational procedures in a given process system. The result of this P-HAZID analysis is given in the form of a spreadsheet and it consists of deviations together with their implications and possible (root) causes. A cause is considered to be a root cause if it is a non-measurable failure mode of a system component (which is an elementary part of the system). For example, a leak on a tank is considered as a root cause. A simple example of a P-HAZID table can be found in *Table* 1.

Using the initial set of differences (deviations) between the characteristic trace and the nominal trace, the set of possible root causes can be found using simple reasoning. For details, refer to *Ref.* [7].

The diagnostic algorithm uses this technique first to find possible P-HAZID row(s) to start from (using the set of initial deviations). Then, following the deviation chains defined by these rows, the algorithm proceeds towards a possible root cause by traversing new rows based on the initial set of deviations. Using this procedure, it may end up at a root cause or at a row with deviations from which it cannot proceed forward, because they are not contained in the initial set of deviations. The algorithm assumes that the root causes are static and they happened before the execution of the procedure began.

Component based diagnosis

It is widely known that complex systems can often be decomposed in a hierarchical way using simple non-dividable elements that are called components. The connection of these components is usually specified in terms of a graph called flowsheet. Such a decomposition can be used effectively for the operation of a reasoning-based diagnostic algorithm.

Component based structural decomposition

The above mentioned fault diagnosis based on the P-HAZID analysis is only developed for process systems

Table 1: A simple example of a P-HAZID table. Inputs: op=**open**, cl=**closed**. Outputs: 0=**no**, L=**low**, N=**normal**. Deviations: NH=**never-happened**, LAT=**later**, EAR=**eariler**, SML=**smaller** and GRE=**greater**. Faults: **TANK-LEAK** is the leak of the tank and **POS-BIAS** is the positive bias failure of the tank level sensor.

Cause	Deviation	Implication
TANK-LEAK	NH(2;op,cl;L)	NH(3;op,cl;N)
NH(2;op,cl;L)	NH(3;op,cl;N)	NH(4;op,op;N)
TANK-LEAK	SML(2;op,cl;L)	SML(3;op,cl;N)
SML(2;op,cl;L)	SML(3;op,cl;N)	SML(4;op,op;N)
POS-BIAS	GRE(1;op,cl;0)	GRE(2;op,cl;L)
GRE(1;op,cl;0)	GRE(2;op,cl;L)	GRE(3;op,cl;N)
GRE(2;op,cl;L)	GRE(3;op,cl;N)	NH(4;op,op;N)
POS-BIAS	NH(1;op,cl;0)	EAR(2;op,cl;L)
NH(1;op,cl;0)	EAR(2;op,cl;L)	EAR(3;op,cl;N)

Figure 1: Process system consisting of 3 similar tanks

consisting of different individual components in [7]; the possible redundancy of such systems (e.g. multiple components of the same characteristics) were not taken into account. However, complex process systems in practice can be decomposed into a connected network of more simple components. For example, the process system in *Fig.* 1 can be decomposed into three smaller similar components each formed by an input and an output valve and a tank.

When developing the decomposed form of a process system it can happen that some elements are part of multiple subsystems as in the case of valves VB and VC in *Fig.* 1. These elements are called boundary components, and are assumed to be error-free during the diagnosis.

Traces affecting different components can also be decomposed into a chain of trace fragments each referring to a single component of the trace. Events in such a trace fragment have only a subset of inputs and outputs of the united trace (only the inputs and outputs of the particular component that is present in them). Fragments also have information about the next trace fragment (called the next trace), and there is a starting condition (an event) that is associated with them to help the diagnosis. Along with the trace fragment, each component has its own associated P-HAZID spreadsheet.

The component based diagnostic algorithm

Applying the diagnostic approach described in [7] on a decomposed process system, the components can be diagnosed separately against faults, treating them as a whole system during diagnosis. After the separate diagnosis, the root causes can be collected and the resulting set of root causes yields to the set of root causes in the united system. Using the component decomposition, the size of the HAZID information required can be made lower in cases when similar connected subsystems form the process system to be diagnosed. On the other hand,
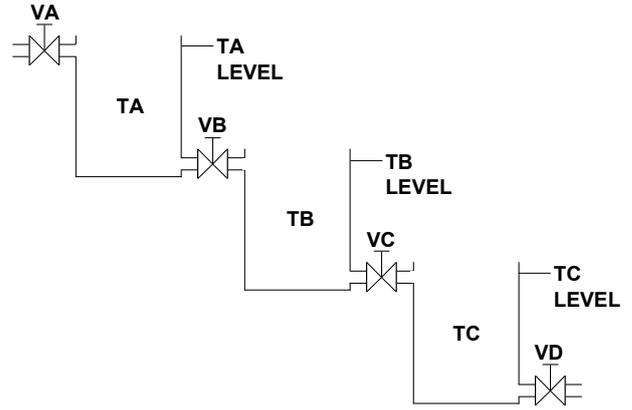
system-level deviations need to be converted into component deviations by aligning their time and reducing the inputs and outputs in their associated events to component-level inputs and outputs.

The diagnosis is then performed by comparing the whole nominal trace with the characteristic trace, and then distributing the deviations (differences) among the components. Before distributing, time alignment and reduction of input and output states to component level are performed (including the component's boundary elements). After the distribution, component-level diagnosis may begin to explore possible (root) causes on the component-level.

The diagnosis proceeds along the flow direction by starting from the first component, takes the deviations, generates the set of possible root causes from them, and then tries to proceed to the next component by checking the starting condition of the next trace fragment - if there is any. If the start condition is fulfilled, the diagnosis continues, otherwise it halts. For example, in the case of *Fig.* 1 the second fragment might have a start condition containing a statement about the minimum level of fluid in tank TA, and in the case of the congestion of valve VA no fluid is coming into a system, therefore, tank TA is not even filling up to the specified minimum level. In this case the diagnosis stops. The result of the diagnostic algorithm is always the union of identified and non-identified root causes created by the consecutive diagnostic algorithm that runs on the components of the consequent nominal trace fragments.

For reference, the whole diagnostic algorithm is presented as a pseudo-code in Algorithm 1. The algorithm collects all root causes (sets **INC** and **IRC**) given a component decomposition, a starting component, and a possibly faulty characteristic trace.

Case study

In the case study the diagnosis procedure for the simple process system in *Fig.* 1 containing three sequentially connected tanks is used. The measured output of the tanks

---

Algorithm 1 Component-based reasoning procedure

1: $\mathbf{INC} \leftarrow \{\emptyset\}$
2: $\mathbf{IRC} \leftarrow \{\emptyset\}$
3: $actualComponent \leftarrow startComponent$
4: $continue \leftarrow \mathbf{true}$
5: $shift \leftarrow 0$
6: **while** $continue$ **do**
7:     $\mathbf{DEV} \leftarrow \text{GENERATEDEVIATIONS}(actualComponent,\ chrTrace,\ shift)$
8:     $\mathbf{FDP} \leftarrow \text{COLLECTFINALDEVIATIONPAIRS}(\mathbf{DEV})$
9:     **for all** $pair \in \mathbf{FDP}$ **do**
10:         $startDeviation \leftarrow proj_1(pair)$
11:         $startImplication \leftarrow proj_2(pair)$
12:         $\text{REASON}(startDeviation, startImplication, actualComponent.phazid)$
13:     **end for**
14:     **if** $actualComponent$ **has successive component and its start condition evaluates to true then**
15:         $shift \leftarrow length(actualComponent.trace) - 1 + shift$
16:         $actualComponent \leftarrow \text{GETCOMPONENT}(actualComponent.successiveComponent)$
17:     **else**
18:         $continue \leftarrow \mathbf{false}$
19:     **end if**
20: **end while**
21: **function** GENERATEDEVIATIONS(component, chrTrace, shift)
22:     $\mathbf{DEV} \leftarrow \{\emptyset\}$
23:     $nomTrace \leftarrow component.trace$
24:     $reducedTrace \leftarrow \text{TRIMTRACE}(chrTrace, shift, shift + component.trace.length)$
25:     $locChrTrace \leftarrow \text{CONVERTTOCOMPONENTLEVEL}(reducedTrace)$
26:     **for** $T := 1$ to $length(nomTrace)$ **do**
27:         **for all** deviation D of $locChrTrace$ from $nomTrace$ at time T **do**
28:             $\mathbf{DEV} \leftarrow \mathbf{DEV} \bigcup(D)$
29:         **end for**
30:     **end for**
31:     **return DEV**
32: **end function**
33: **procedure** REASON(deviation, implication, phazid)
34:     **if** $\exists R \in \text{ROWS}(phazid), deviation = dev_{phazid}(R), implication = imp_{phazid}(R)$ **then**
35:         **for all** $\{R, dev_{phazid}(R) = deviation$ **and** $imp_{phazid}(R) = implication\}$ **do**
36:             **if** $cause_{phazid}(R) \in \mathbf{RC}$ **then**
37:                 $\mathbf{IRC} \leftarrow \mathbf{IRC} \bigcup cause_{phazid}(R)$
38:                 **return**
39:             **else if** $cause_{phazid}(R) \in \mathbf{DEV}$ **and** $cause_{phazid}(R) \prec dev_{phazid}(R)$ *in* $\mathbf{DEV}$ **then**
40:                 $\text{REASON}(cause_{phazid}(R),\ dev_{phazid}(R),\ phazid)$
41:             **else**
42:                 $\mathbf{INC} \leftarrow \mathbf{INC} \bigcup cause_{phazid}(R)$
43:                 **return**
44:             **end if**
45:         **end for**
46:     **else**
47:         $\mathbf{INC} \leftarrow \mathbf{INC} \bigcup cause_{phazid}(R)$
48:         **return**
49:     **end if**
50: **end procedure**

Table 2: Tank fill operational procedure.

Time	Input values				Output values		
	VA	VB	VC	VD	TA	TB	TC
1	op	cl	cl	cl	0	0	0
2	op	cl	cl	cl	L	0	0
3	op	op	cl	cl	N	0	0
4	op	op	cl	cl	N	L	0
5	op	op	op	cl	N	N	0
6	op	op	op	cl	N	N	L
7	op	op	op	op	N	N	N

Table 3: Normal fill in a single tank with no faults.

Input valve	Output Valve	Tank Level
op	cl	0
op	cl	L
op	op	N

is the tank level that takes its values from Q_e in *Eq.* (1), and the input variables are the valve positions (**open** (op) or **closed** (cl)).

Components, operational procedure, and nominal trace

Every tank may contain no fluid (the tank level is equal to **no** (0)), may be low on fluid (level value is **low** (L)), or might have normal fluid level (level value is **normal** (N)). In every time instant the level increases by one qualitative magnitude (i.e. from **no** to **low** or from **low** to **normal**) if fluid is coming through the input valve but the output valve is closed. Due to the same size of the valves the effect of fluid flow out of the system is similar, but in the opposite direction (from **normal** to **low** or from **low** to **no**). The valve positions (**open** (op) or **closed** (cl)) can be changed by the operator; they are considered as inputs of the system. Leak in the tank is assumed to be equal to the size of an open valve (i.e. a quite substantial leak).

The considered operational procedure is the initial filling of all the three tanks with fluid (the fill operational procedure in short), and is described in detail in *Table* 2.

The process system can be decomposed into three components, therefore the fill operational procedure can also be partitioned into three identical procedure fragments along the component boundaries (the VB and VC valves).

The fragment abstracted from the three identical trace fragments associated to the three tanks can be observed in *Table* 3. It has only the subset of inputs and outputs which are directly related to the particular tank component - the input and output valve and the tank level.

The corresponding component P-HAZID table can be found in *Table* 4 with some of the component faults and deviations associated to them. Instances of this P-HAZID table are used in the case of all three tanks during diagnosis.

Table 4: P-HAZID table of a single tank component with two valves for a reference trace of *Table* 3. Faults: **TANK-LEAK** is leak of the tank, **POS-BIAS** is the positive bias fault of the level sensor and **NEG-BIAS** is the negative bias failure of the level sensor.

Cause	Deviation	Implication
TANK-LEAK	NH(2;op,cl;L)	NH(3;op,op;N)
TANK-LEAK	SML(2;op,cl,L)	SML(3;op,op;N)
NEG-BIAS	LAT(1;op,cl;0)	NH(2;op,cl;L)
LAT(1;op,cl;0)	NH(2;op,cl;L)	NH(3;op,op;N)
NEG-BIAS	SML(1;op,cl;0)	SML(2;op,cl;L)
SML(1;op,cl;0)	SML(2;op,cl;L)	SML(3;op,op;N)
POS-BIAS	NH(1;op,cl;0)	EAR(2;op,cl;L)
NH(1;op,cl;0)	EAR(2;op,cl;L)	NH(3;op,op;N)
POS-BIAS	GRE(1;op,cl;0)	GRE(2;op,cl;L)
GRE(1;op,cl;0)	GRE(2;op,cl;L)	GRE(3;op,op;N)

Table 5: Tank fill operational procedure with a leak in the second tank TB. The leak caused two different events in the operational procedure related to TB (in **bold**), these differences resulted in the four deviations the diagnosis could start from.

Time	Input values				Output values		
	VA	VB	VC	VD	TA	TB	TC
1	op	cl	cl	cl	0	0	0
2	op	cl	cl	cl	L	0	0
3	op	op	cl	cl	N	0	0
4	**op**	**op**	**cl**	**cl**	**N**	**0**	**0**
5	**op**	**op**	**op**	**cl**	**N**	**0**	**0**
6	op	op	op	cl	N	0	0
7	op	op	op	op	N	0	0

The operation of the diagnostic algorithm

The operation of the diagnostic algorithm is illustrated with the case when a rupture of the second tank is present as a root case (fault).

A characteristic trace with the leak of the second tank can be seen in *Table* 5. The size of the leak is assumed to be larger or equal to the size of an outbound pipe, therefore, the tank cannot fill up and no fluid can flow to the third tank TC.

The starting condition of the TB and TC tank components is the appropriate "normal" level in the preceding tank. In that way it is ensured that diagnosis is done on the operational components only.

The diagnosis of this faulty scenario begins by starting with the first tank component TA. There are no differences (and therefore no deviations) regarding this component. The start condition of the second component is fulfilled, therefore, the diagnosis moves towards the next component TB.

However, in the case of TB the following deviations are found after comparing the nominal and characteristic traces (due to that two consequent events did not hap-

pened, instead, two events happened with lower output values at time instant 4 and 5 in the operational procedure):

- **never-happened**(4;**open**,**closed**;**low**)
- **never-happened**(5;**open**,**open**;**normal**)
- **smaller**(4;**open**,**closed**;**low**)
- **smaller**(5;**open**,**open**;**normal**)

The time instant of the deviations are shifted back by 2 units because the second component's first event happens at the third system-level time instant. After that, the diagnosis is initiated on the HAZID table. Searching for the already found deviations and connecting them to possible root causes (as in the case of the original diagnostic idea in [7]) the leak of the second tank can be found.

Because of the lack of fluid in the second tank, the start condition of the third component is not fulfilled, therefore, the diagnostic process halts at this step resulting in a single possible root cause being the **TANK-LEAK**.

Conclusions

A novel component based extension of the single component diagnostic algorithm presented in [7] is described in this paper. Using the extension, the domain of application can be extended to more complex composite process systems. Driven by the decomposition of the overall system into components, the P-HAZID tables used for diagnosis are processed at component level by the diagnostic algorithm. The extended method is efficient in the cases when the overall process system consists of similar small components.

The component-based diagnostic procedure was described on a formal level, along with its proposed pseudocode. A case study for a process system of multiple components and a simple failure was also provided.

The following improvements are planned to extend the component-based diagnostic approach:

- The procedure is based on the assumption that the boundary elements between different components are free of failures. As a future work, this limitation might be removed by using a higher level reasoning above the components (as in the form of a system-level HAZID table, for example).
- At the moment, the algorithm is only working for already coded static event information in order to find faults in the system. Diagnosis would be more valuable if events could be processed dynamically, thus

the diagnostic procedure could be executed real-time along with the operational procedures.
- Diagnosis would be more accurate if the derivatives of internal states (e.g. the derivative of the tank level) were present in the events.

Acknowledgements

REFERENCES

[1] CAMERON, I. T., RAMAN, R. Process Systems Risk Management. Vol. 6 of Process Systems Engineering (Elsevier Academic Press, San Diego, CA) 2005

[2] AS IEC 61882-2003: Hazard and operability studies (HAZOP studies) Application Guide

[3] VENKATASUBRAMANIAN, V., RENGASWAMY, R., YIN, K., KAVURI, S. N. A review of process fault detection and diagnosis Part I: Qualitative model-based methods, Computers and Chemical Engineering, *2003*, 27(3), 293–311

[4] VENKATASUBRAMANIAN, V., ZHAO, J. S., VISWANATHAN, S. Intelligent systems for HAZOP analysis of complex process plants, Computers and Chemical Engineering, *2000*, 24(9-10), 2291–2302

[5] SELIGMANN, B. J., NÉMETH, E., HANGOS, K. M., CAMERON, I. T. A blended hazard identification methodology to support process diagnosis, Journal of Loss Prevention in the Process Industries, *2012*, 25 , 746–759

[6] NÉMETH, E., LAKNER, R., CAMERON, I., HANGOS, K. M. Fault diagnosis based on hazard identification results, in Preprints of the 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes, Barcelona, Spain, June 30 - July 3, pp. 1515–1520

[7] TÓTH, A., HANGOS, K. M. , WERNER-STARK, A. HAZID information based operational procedure diagnosis method, in 12th International PhD Workshop on Systems and Control, Veszprém, 2012. aug. 27 (A. MAGYAR, ed.), ISBN 978-615-5044-71-7 (University of Pannonia, Veszprem), pp. 1–6 (on CD)

ROLE OF STEADY STATE DATA RECONCILIATION
IN PROCESS MODEL DEVELOPMENT

Barbara Farsang[✉1], Sándor Németh[1], and János Abonyi[1]

[1]University of Pannonia, Department of Process Engineering, H-8200 Veszprém, Egyetem Street 10., HUNGARY
[✉]E-mail: farsangb@fmt.uni-pannon.hu

In chemical and hydrocarbon industry operational efficiency is improved by model-based solutions. Historical process data plays an important role in the identification and verification of models utilized by these tools. Since most of the used information are measured values, they are affected by errors influencing the quality of these models. Data reconciliaton aims the reduction of random errors to enhance the quality of data used for model development resulting in more reliable process simulators. This concept is applied to the development and validation of the complex process model and simulator of an industrial hydrogenation system. The results show the applicability of the proposed scheme in industrial environment.

Keywords: data reconciliation; flowsheeting simulator; validation; balance equation; industrial hydrogenation system

Introduction

In recent years, development of industrial technologies has been determined by the proliferation of computing and information technology. Today's technologies are characterized by widespread application of process engineering tools. The success of these tasks depends on the accuracy of data and applied models [1]. Thanks to the evolution of information technology, on-line and historical process data — coming from chemical process systems — are available. The collected data provide the opportunity for engineers to better understand the processes, anomalies, and malfunctions [2]. Monitoring of process variables allows us to ensure the consistent product quality. Collected data can also be used for the development and validation of process simulators. However, measurements are always affected by errors during the measurement, processing and transmission of the measured signal. Errors (gross, random, bias) in measured data affect the quality of process models and can lead to significant decrease in plant performance. Estimation of true conditions of process states is important to achieve optimal process monitoring, control and optimization.

Therefore, several methods have been developed to minimize measurement errors thereby enhancing the reliability, accuracy and precision of data. First, analogue and digital filters were used to reduce the effect of high frequency noise [3]. Gross errors were detected with various data validation methods, which include checking whether the measured data and the rate at which it is changing is within predefined operational limits. Nowadays, smart sensors are used for determine whether there is any hardware problem or the measured

data are appropriate. New methods were statistical quality control tests that are applied to each measured variable separately. Although these methods improve the reliability and accuracy of the measured data, they do not ensure consistency of the data with respect to the inter-relationships between different process variables. Therefore new methods – data reconciliation and gross error detection – are developed in chemical engineering [4]. The main difference between data reconciliation and other filtering methods is that data reconciliation uses process model constraints and that the results satisfy constraints and balance equations. Data reconciliation techniques can be applied to reduce random errors of measurements, while the other techniques mentioned above do not. Gross error detection can be used for eliminate systematic errors so simultaneous data reconciliation and gross error detection have emerged as a key of online optimization [5] (see *Fig. 1*).

Data reconciliation techniques take minimal corrections of the measured variables to satisfy a set of model constraints. Based on the difference between the measured and reconciled data, the following questions can be answered. Can we consider a related set of measurements acceptable based on our previous knowledge of the system? Are the measurements consistent? If not, what can be the source of the error? Based on the available measurements and a prior knowledge, what is the most likely state of the system? Answers to these questions are important in development of technology (e.g., monitoring, optimization, simulation, control, instrument maintenance). Using this technique, we can verify the acceptability of measurements, improve the accuracy of measurements, estimate model parameters and unmeasured variables, and it can be used
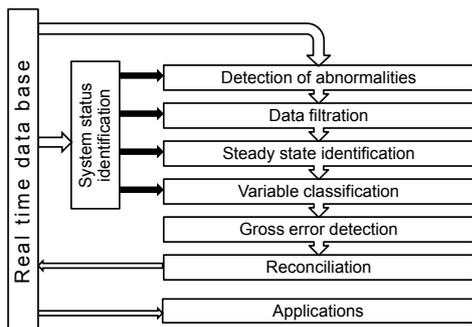
Figure 1: General methodology for on-line data correction [6]

Figure 2: Types of data reconciliation [7]

for fault detection.

In order to ensure the consistency of measured process variables, models are used in data reconciliation. During data reconciliation, information is obtained from both measurements and process models. Depending on the types of the models, data reconciliation techniques can be separated into two different problems: steady state and dynamic (see *Fig. 2*). In the case of steady state data reconciliation, model constraints are algebraic equations. If we are dealing with dynamic processes, differential equations are used as model constraints. In the case of linear data reconciliation, model constraints are defined by linear equations. The simplest data reconciliation technique is the steady state linear method that can describe, e.g., the mass balance of technology. Moreover, most chemical processes have nonlinear characteristics and constraints are nonlinear equations; this problem can be called as nonlinear data reconciliation.

The goal of the presented research is to propose a methodology to support the development and validation of complex process models and simulators based on increasing the quality of measurement data used for (kinetic) parameter identification and validation. In section *Literature survey*, a historical background of steady data data reconciliation is briefly reviewed. The basic method of linear and nonlinear steady state data reconciliation is described in section *Theoretical background of steady state data reconciliation*. Data reconciliation cannot be used in process development only, but it is a useful tool in model improvement too. The principle of this application is introduced in section *Combined application of data reconciliation and flowsheeting simulator*. Our approach aims the reduction of random errors to enhance the quality of data using data reconciliation and flowsheeting simulator simultaneously. The method is described in subsection *The developed model based data reconciliation technique*. The proposed approach is illustrated based on an industrial hydrogenation system.
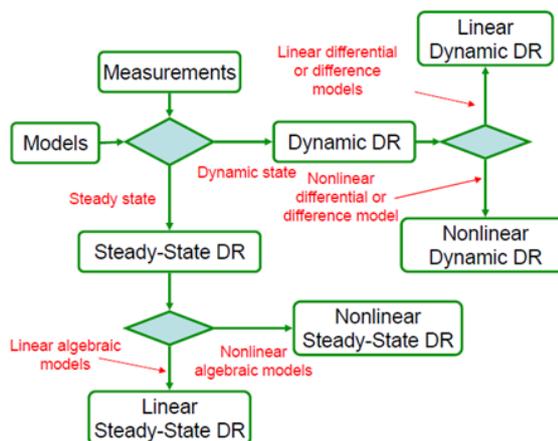
In section *Case study*, the analyzed technology is introduced and the applicability of the method is illustrated on the basis of the mass balance of the hydrogenation system. Then, some consequences are drawn in section *Conclusion*.

Literature survey

KUHAN and DAVIDSON presented the first study about minimizing measurement errors in chemical process industry using a data reconciliation technique in 1961 [4]. They described the steady state linear data reconciliation technique and demonstrated a general solution for mass balance of technology. This seminal paper has started a new research area in process engineering. NOGITA [8] and MAH [9] improved a new method capable of detection of gross errors. Later, MAH [10] found that the base method requires a process system in steady state conditions and that the process constraints must be linear. CAO and RHINEHART confirmed that the methodology will be successful if process data come from steady state conditions [11]. In these cases, data reconciliation problems can be solved with standard numerical methods like sequential quadratic procedures [12] and quasi-Newton algorithms [13]. A multiple Gauss-Jordan elimination algorithm was published by MADRON and VEVERKA [14]. The classification method is based on matrix decomposition and permutation. The linear method was improved by CROWE (1986) who published a new technique to solve nonlinear data reconciliation using matrix projection [15]. Furthermore, BAGAJEWICZ and JIANG [16] presented an improved method that is applicable for dynamics system. The developed method from PRATA [17] is suitable for examination of dynamic nonlinear process systems.

Despite of the numerous good reviews of data reconciliation techniques, industrial application of this method started only in the 1980s. There are several complex software packages supporting data

Table 1: Industrial application of data reconciliation (newer industrial scenario and publications are added to PRATA's summary [18])

Industrial scenario	Author (year)
Absorption refrigeration systems	MARADIAGA et al. (2013) [19]
Beverage alcohol distillation plant	MEYER et al. (1993), NOUNOU and BAKSHI (1999), SCHLADT and HU (2007)
Chemical extraction plant	HOLLY et al. (1989)
Ethylene and ammonia plant	SANCHEZ et al. (1992), SANCHEZ and ROMAGNOLI (1996), PLÁCIDO and LOUREIRO (1998)
Exxon chemical process	MCBRAYER et al. (1998) , SODERSTROM et al. (2000)
Gas pipeline systems	BAGAJEWICZ and CABRERA (2003)
Gases network in an iron and steel making plant	YI and HAN (2004)
Hydrogen plant	BUSSANI et al. (1995), CHIARI (1997), SARABIA et al. (2012) [20]
Industrial coke-oven-gas purification process	FABER et al. (2006) HU and SHAO (2006)
Industrial distillation column	ISLAM (1994), WEISS et al. (1996), SANCHEZ et al. (1996), BOUROUIS et al. (1998), LI et al. (2001), BHAT and SARAF (2004), CHATTERJEE and SARAF (2004), CHEN et al. (2013) [21], KELLER et al. (2012) [22]
Industrial ETBE reactor	DOMINGUES et al. (2012) [23]
Industrial furnace	PIERUCCI et al. (1996), EKSTEEN et al. (2002)
Industrial hydrometallurgical plants for a gold extraction	DE ANDRADELIMA (2006)
Industrial polymerization reactor	VIEIRA et al. (2003), PRATA et al. (2006, 2008, 2009, 2010 [24])
Industrial synthesis gas for production of ammonia	CHRISTIANSEN et al. (1997)
Industrial utility plant	LEE et al. (1998)
Methyl-terc-butyl-ether plant	AL-ARFAJ (2006)
Mineral and metallurgical plants	VASEBI et al(2012) [25]
Nuclear power reactor	VALDETARO (2011) [26]
Refinery	PICCOLO and DOUGLAS (1996), ZHANG et al. (2001)
Sulfuric acid plant	CHEN et al. (1998), OZYURT and PIKE (2004)
Turbine cycle of a boiling water reactor	SUNDE and BERG (2003)
Vinyl acetate and ketene plants	DEMPF and LIST (1998)
Water processes	MARTINS et al. (2010) [27]

reconciliation (for instance SigmaïňĄne, DataCon, Vali, Inlibra). These tools have interfaces to information systems used to manage process data and have special model building functionalities that allow almost automatic building of plant-wide balance equations. Nowadays, data reconciliation is widely applied in various processing industries (e.g. refining, chemical industry, metals, mining, and power industry). Despite of this fact, relatively few articles deal with industrial applications (PRATA et al. prepared a summary in 2009 [18]; newer applications and papers are collected in *Table 1*).

STANLEY et al. classified variables as observable and unobservable [28]. CROWE categorized the variables from another point of view: the basis of classification is the given measured variable can be calculated from other measured variables using process models or not. The names of the two classes are redundant and non-redundant measurements [29]. ALI and NARASIMHAN applied graph theory to analyse the sensor network, classify the variables, and compute the redundancy degree of each variable. They claim that not only measured variables can be redundant. Those unmeasured variables that can be estimated in multiple ways are also termed as redundant [30].

Theoretical background of steady state data reconciliation

In this section, the basis of linear and nonlinear data reconciliation techniques are described. Different values are distinguished. Real values (without error) of variables are designated by diacritic caron (ˇ), measured values by tilde (˜), while estimated values by circumflex (ˆ).

In general, optimal estimates for process variables by data reconciliation are solutions to a constrained least-squares or maximum likelihood objective function, where measurement errors are minimized with process model constraints. The steady state data reconciliation problem can be formulated as an optimization problem by minimizing

$$J(\check{y}, \check{z}) = (\tilde{\mathbf{y}} - \check{\mathbf{y}})^T \mathbf{V}_{\tilde{\mathbf{d}}}^{-1} (\tilde{\mathbf{y}} - \check{\mathbf{y}}) \tag{1}$$

subject to

$$\mathbf{f}(\check{y}, \check{z}) = 0 \tag{2}$$

and

$$\mathbf{g}(\check{y}, \check{z}) \geq 0. \tag{3}$$

The solution of the optimization problem is performed with the following simplifying assumptions

- The measurement error is independent from the balance variables.
- The expected value of measurement error is zero.
- The measurement error is in a normal (Gauss) distribution.
- Errors of measurements are independent from each other (diagonal covariance matrix).

- The covariance matrix is positive definite, so its inverse matrix exists.

In the following two subsections, solution methods are presented in the cases of linear and nonlinear models.

Linear data reconciliation

In the linear case, when all variables are measured, *Eq. (1)* means minimizing

$$J(\check{y}, \check{z}) = (\tilde{\mathbf{y}} - \check{\mathbf{y}})^T \mathbf{V}_{\tilde{\mathbf{d}}}^{-1} (\tilde{\mathbf{y}} - \check{\mathbf{y}})$$

subject to

$$\mathbf{A}\check{\mathbf{y}} = 0. \tag{4}$$

Gauss distribution $(\tilde{\mathbf{d}} \sim N(0, \mathbf{V}_{\tilde{\mathbf{d}}}))$ is assumed, so the density function of measurement error is

$$f\left(\tilde{\mathbf{d}}\right) = C \cdot \exp\left[-0.5 \cdot \tilde{\mathbf{d}}^T \mathbf{V}_{\tilde{\mathbf{d}}}^{-1} \tilde{\mathbf{d}}\right] \tag{5}$$

from which it follows that

$$f\left(\tilde{\mathbf{y}}\right) = C \cdot \exp\left[-0.5 \cdot (\tilde{\mathbf{y}} - \check{\mathbf{y}})^T \mathbf{V}_{\tilde{\mathbf{d}}}^{-1} (\tilde{\mathbf{y}} - \check{\mathbf{y}})\right] \tag{6}$$

and that the

$$\check{\mathbf{f}} = \mathbf{A}\check{\mathbf{y}} - \mathbf{b} = 0 \tag{7}$$

criterion is satisfied in accordance with the balance equation. The essence of the most likely estimation: if $\tilde{y}$ is given and its density function is known, which $\tilde{y} \to \hat{y}$ parameter of density function will be $\mathbf{f}$ (function vector of equality model constraints).

Since the logarithm function is strictly monotonously increasing, the maximum of the logarithm of the $f(y)$ function recording only positive values is at the same place where the maximum of the argument is. Exploiting it on the above function and using the Lagrange multipliers method, the result is a vector equation system with two unknowns:

$$\mathbf{V}_{\tilde{\mathbf{d}}}^{-1} \hat{\mathbf{y}} - \mathbf{A}^T \hat{\lambda} = \mathbf{V}_{\tilde{\mathbf{d}}}^{-1} \tilde{\mathbf{y}} \tag{8}$$

$$\mathbf{A}\hat{\mathbf{y}} = \mathbf{b} \tag{9}$$

The solution of the vector equation system is

$$\begin{aligned}
\hat{\mathbf{y}} &= \left(\mathbf{I} - \mathbf{V}_{\tilde{\mathbf{d}}}\mathbf{A}^T \left(\mathbf{A}\mathbf{V}_{\tilde{\mathbf{d}}}\mathbf{A}^T\right)^{-1}\mathbf{A}\right)\tilde{\mathbf{y}} \\
&+ \mathbf{V}_{\tilde{\mathbf{d}}}\mathbf{A}^T \left(\mathbf{A}\mathbf{V}_{\tilde{\mathbf{d}}}\mathbf{A}^T\right)^{-1}\mathbf{b}
\end{aligned} \tag{10}$$

Thus, the estimation requires a matrix-vector multiplication and a vector addition. Because $\tilde{\mathbf{y}}$ is only the variable, in the other operations there are only constants [31].

In practice, not all streams and properties are measured in plant due to physical and economical reasons. The data reconciliation technique is suitable for estimating unmeasured variables. In this case, the optimization problem can be solved by the method of

Projection Matrix. In the first step, the incidence matrix is separated into two parts:

$$\mathbf{A}_y \check{\mathbf{y}} + \mathbf{A}_z \check{\mathbf{z}} = 0. \tag{11}$$

Unmeasured flows ($\check{z}$) should be eliminated by pre-multiplying both sides by a projection matrix, such as $\mathbf{PA}_z\check{\mathbf{z}} = 0$. *Eq. (4)* then can be rewritten as

$$\min J(\check{y}, \check{z}) = (\tilde{\mathbf{y}} - \check{\mathbf{y}})^T \mathbf{V}_{\tilde{\mathbf{d}}}^{-1} (\tilde{\mathbf{y}} - \check{\mathbf{y}}) \tag{12}$$

subject to

$$\mathbf{PA}_y \check{\mathbf{y}} = 0. \tag{13}$$

The projection matrix is from Q-R factorization of matrix $\mathbf{A}_z$:

$$\mathbf{A}_z = \mathbf{QR} = [\mathbf{Q}_1\ \mathbf{Q}_2]\, [\mathbf{R}_1\ \mathbf{0}]^T \tag{14}$$

The projection matrix is the transformation of $\mathbf{Q}_2$ matrix ($\mathbf{P} = \mathbf{Q}_2^T$). If $\mathbf{A}$ matrix is replaced by matrix $\mathbf{PA}_y$ in *Eq. (10)*, then

$$\hat{\mathbf{y}} = \left(\mathbf{I} - \mathbf{V}_{\tilde{\mathbf{d}}}(\mathbf{PA})^T \left((\mathbf{PA}) \mathbf{V}_{\tilde{\mathbf{d}}}(\mathbf{PA})^T \right)^{-1} (\mathbf{PA}) \right) \tilde{\mathbf{y}}$$
$$+ \ \mathbf{V}_{\tilde{\mathbf{d}}}(\mathbf{PA})^T \left((\mathbf{PA}) \mathbf{V}_{\tilde{\mathbf{d}}}(\mathbf{PA})^T \right)^{-1} \mathbf{b}. \tag{15}$$

After reconciled values of measured variables are obtained, the next step is to estimate the unmeasured variables using the information from process models (*Eq. (11)*) as

$$\mathbf{A}_y \check{\mathbf{y}} = -\mathbf{A}_z \check{\mathbf{z}} \tag{16}$$

This linear equation can be solved to get estimated values of the unmeasured variables. Usually, the number of equations is greater than the number of unmeasured variables. The least-squares technique can be applied and the solution is

$$\hat{\mathbf{z}} = -(\mathbf{A}_z^T \mathbf{A}_z)^{-1} \mathbf{A}_z^T (\mathbf{A}_y \hat{\mathbf{y}}). \tag{17}$$

Nonlinear data reconciliation

The data reconciliation problem can be extended to nonlinear steady state models. It is necessary because processes in chemical industry cannot be described only by linear models; behaviour of chemical processes are often nonlinear. If we wish to simultaneously reconcile mass flow, composition, or temperature measurements, mass, component, or energy balances have to be included as constraints. Sometimes, inequality constraints have to be defined (e.g. mass/mole fraction of components have to be in the [0:1] interval). Moreover, if we take thermodynamic equilibrium relationships and complex correlations for thermodynamic and physical properties as constraints, nonlinear data reconciliation techniques have to be used.

If all variables are measured, the bilinear data reconciliation problem can be reduced to a linear problem by introducing the "measured" compound flows ('component flows' or 'energy flows'). Therefore, solution of the nonlinear data reconciliation problem is
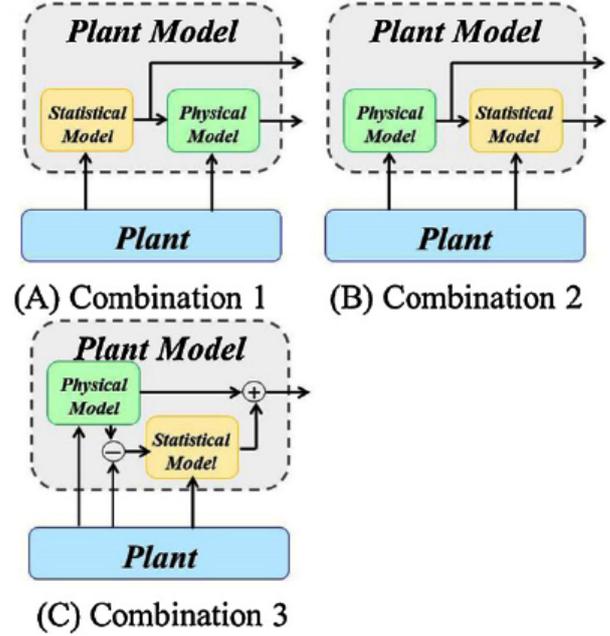
Figure 3: Integrated application of statistical and white box models [32]

presented on the basis of the situation when measured and unmeasured variables occur.

The objective function is the same as in the previous case (*Eq. (1)*). The difference is that constraints are nonlinear equations. This problem can be reduced to the linear case. This method is referred to as successive linearisation: nonlinear constraints can be linearised around working points using first-order Taylor series. The solution of the linearisation can be written as

$$\mathbf{h} = \mathbf{B}_y \check{\mathbf{y}} + \mathbf{B}_z \check{\mathbf{z}} - \mathbf{f}(\check{\mathbf{y}}_i, \check{\mathbf{z}}_i). \tag{18}$$

The initial values are usually the raw values. After linearisation the procedures of data reconciliation is described in subsection Linear data reconciliation (Q-R factorization, reconciliation of measured values, then estimation of unmeasured variables). If the reconciled and initial values are far from each other, new iteration begins. The initial value will be the result of the previous iteration. Different criteria can be defined: number of iteration or $\|\hat{\mathbf{y}}_n - \hat{\mathbf{y}}_{n-1}\|$ is smaller than the specified tolerance values.

Successive linearisation is a relatively simple and fast solution, but variable bounds cannot be handled with this method. Another option to solve nonlinear data reconciliation problem is using Nonlinear Programming (NLP) techniques that can estimate measured and unmeasured variables simultaneously. Sequential quadratic programming (SQP) and generalized reduced gradient (GRG) are usual techniques in handling nonlinear problems. These methods are more computationally demanding, but they are numerically more robust [33].
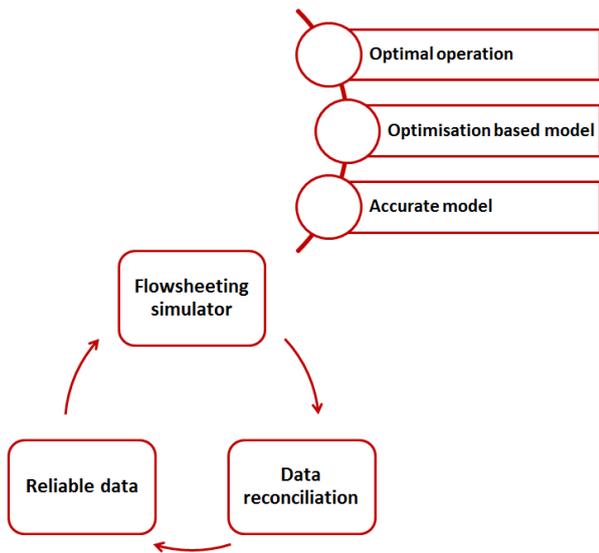
Figure 4: Iterative model development

Figure 5: Options of data processing in flowsheeting simulator

Combined application of data reconciliation and flowsheeting simulator

Mathematical models describe the connection between the relevant properties of the analyzed phenomena, process, or activities. In the case of complex chemical processes, mathematical models usually contain differential equations whose solution is not always possible analytically. The process simulator (it is often applied to improve complex industrial processes) contains models of the equipments so their mathematical description is unnecessary for the users. In addition, it includes different thermodynamic models and component databases. The flowsheeting simulator can be defined as "Use of a computer program to quantitatively model characteristic equations of a chemical process" [34]. The simulator is used in batch processing, integrated process engineering (for example, economic analysis and supply chain forecast, supervisory process control, on-line modeling and optimization, safety and reliability analysis), and process synthesis and design (for instance, heat integration, conceptual design). However, lots of information are needed from technology for the simulator to describe the real process exactly. Many of the needed data are measured (e.g., flow, composition, temperature, pressure). If the input of the simulator is faulty, the simulator cannot give reliable results.

A new direction is the combination of simulators and data based statistical models that can handle random and gross errors [23, 35]. The structure of these hybrid models are shown in *Fig. 3*.

- In the first case, the statistical model is the input of the physical model in the form of differential or algebraic equations or a complex flowsheeting simulator. In this case, the statistical model is used to estimate parameters and phenomena that are difficult to model.
- Combination 2 shows the case when outputs of the physical model are transformed by a statistical model.
- In the third option, the difference between measured and calculated variables are the inputs of the statistical model used for correction.
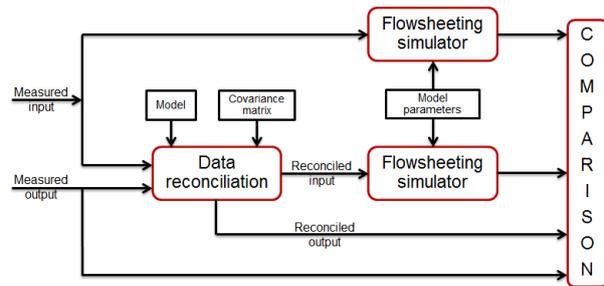
Model-based data reconciliation techniques are similar to the third approach. Using data reconciliation technique random errors can be filtered. When accurate data are available, the flowsheeting simulator gives more reliable results that make further development of models possible. This means that data reconciliation allows us to check the reliability of these measurements [36] and reconsolidated data can be used to build accurate models.

The developed model based data reconciliation technique

Our goal is to design an expert system that can be used to check the acceptability of measurements and improve the flowsheeting simulator of technology. The proposed method is based on model-based data reconciliation techniques. Using data reconciliation, faulty measurements can be found. Random errors can be filtered using discrete Fourier transform. The frequency of the noise is much higher than the frequency of basic process so the discrete Fourier transform gives a graphical view about the differences between the basic process and the noise. Thus, the filtration of the frequency can be determined so the noise can be separated from the basic process. If the amplitude of the noise is known, the standard deviation can be estimated. Thereby, input of simulator is error-free, so systematic model mismatch can be recognized and the tuning of the model can be initiated. Data reconciliation requires an accurate model whose parameters require reconciled process data. In th case of an efficient iterative procedure, a properly working simulator can be developed (see *Fig. 4*).

Since the simulator is able to process historical measured data, the difference between the theoretically achievable values and the measured outputs can give useful information (see *Fig. 5*). Four different output
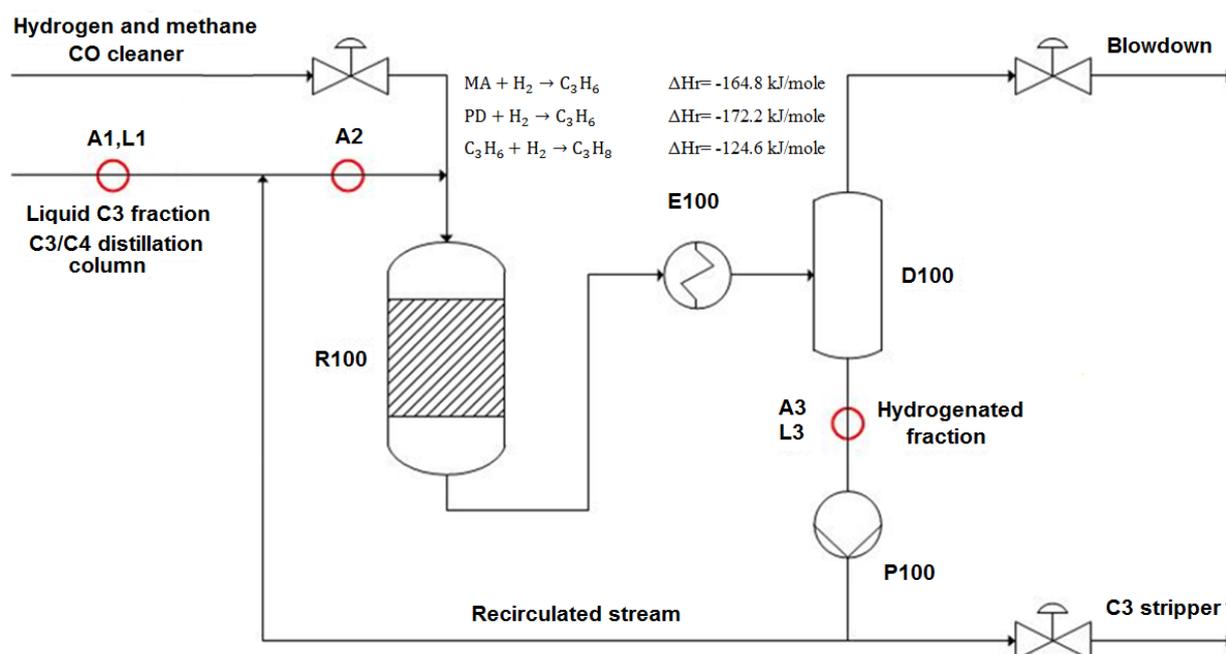
Figure 6: Flowsheet diagram of the C3 hydrogenation system

values can be compared in the case of the same variable. Measured output values are known. If all necessary information is available for data reconciliation, we get the reconciled output value. In addition, if process model is prepared in the flowsheeting simulator, we have two options for analysis: the input of the simulator can be the measured or the reconciled value.

The difference between outputs can be used for two purposes. If the difference is insignificant, the measured data is acceptable and measuring instruments operate properly; calibration is not necessary. Moreover, it helps in the development of the flowsheeting simulator. If the calculated results with reconciled input are far from the reconciled output, the simulator do not describe the real process properly; maybe a parameter is not accurate or it needs structural change.

The proposed method is illustrated on the basis of an industrial hydrogenation system. The case study shows examples for both cases; when the difference indicates the conformity of measurements and when it warns that some model parameters are incorrect.

Case study

The Tisza Chemical Group Plc. (TVK) is the largest petrochemical company of Hungary where polymer raw materials (ethylene, propylene, butylenes, etc.) are produced by steam cracking of naphtha or gasoline. At high temperature, numerous free radical reactions occur. Cracked gas includes light components that are produced in larger amounts, for example, methane, ethane, ethylene, acetylene, propane, propylene, methyl-acetylene, propadiene, n-butane, isobutane,

1,3-butadiene, etc. Some of them are undesirable, because they poison the catalysts for polymerization reactions (e.g. acetylene, methyl-acetylene, propadiene). Another problem is that separation of these components is difficult from the main products by distillation. Due to these problems, undesirable hydrocarbons are hydrogenated. Hydrogenation process of methyl-acetylene and propadiene is presented in this case study (see *Fig. 6*).

The C3-selective hydrogenation process transforms methyl-acetylene and propadiene to propylene by catalytic reaction avoiding the transformation of propylene to propane. The methyl-acetylene and propadiene content of hydrogenated C3 fraction should be less than 1000 ppm. The concentration of the main components (methyl-acetylene, propane) is measured by online analysers (A1, A2 and A3) and there are two places where sample is taken twice a day for laboratory analysis (L1 and L3).

Hydrogen and liquid olefin stream are fed to the reactor where chemical reactions take place. Outlet stream of the reactor is cooled by water so the gaseous C3 component condenses and the hydrogen and the liquid olefin phases are separated. The recirculated stream has two important functions: cools the reactor and dilutes the inlet C3 stream. Blowdown is required due to the accumulation of inert components.

The most frequent type of hydrogenation reactors is the trickle-bed reactor in the olefin plant. Liquid olefin feed and gaseous hydrogen pass through the catalyst bed in the same direction: from top to bottom. Cooling of reactor is provided by vaporization liquid flow. The reactions occur in the top region of the reactor. The
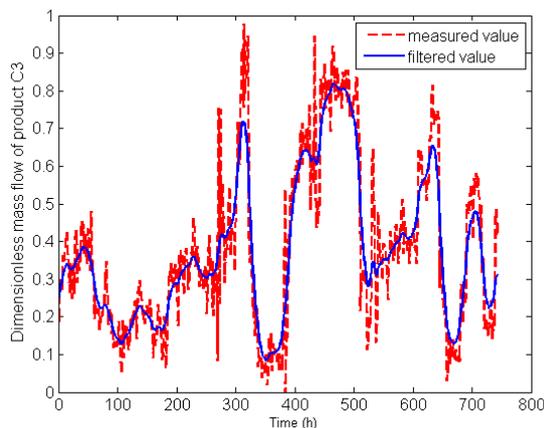
Figure 7: Measured and filtered values of product C3 fraction
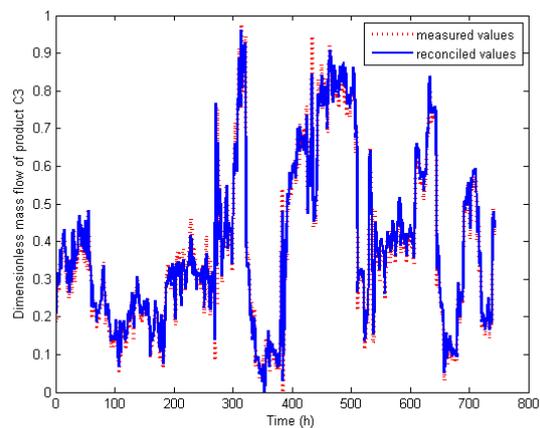
Figure 8: Measured and reconciled values of product C3 fraction

reactions release a large amount of heat. The heat generation is influenced by the methyl-acetylene and propadiene (MAPD) content of the inlet C3 and by the partial pressure of the hydrogen. The MAPD content of the inlet flow of system is high, so the inlet flow of the reactor is diluted. In the range of $10 - 80\,^\circ\mathrm{C}$ temperature, the average reaction heat of main reactions are the following:

$$\mathrm{MA} + \mathrm{H_2} = \mathrm{C_3H_6} \quad \Delta H_\mathrm{r} = -164.8\,\mathrm{kJ/mole} \quad (19)$$

$$\mathrm{PD} + \mathrm{H_2} = \mathrm{C_3H_6} \quad \Delta H_\mathrm{r} = -172.2\,\mathrm{kJ/mole} \quad (20)$$

$$\mathrm{C_3H_6} + \mathrm{H_2} = \mathrm{C_3H_8} \quad \Delta H_\mathrm{r} = -124.6\,\mathrm{kJ/mole} \quad (21)$$
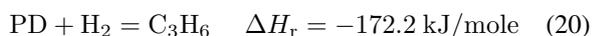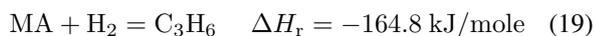
In the following sections, released heat vaporizes a part of the liquid phase. This phenomenon also ensures the cooling of the product flow.

The simulator has been developed in Aspen Plus software as part of an expert system used for the monitoring and qualification of the operation of the technology. Calculations provide useful information about unmeasured variables, validity of on-line analyzers, and efficiency of catalysts. The residence time is small (less than one minute). The sample time was five minutes, so the steady state simulation gives the same result as the dynamic simulation. Historical process data entered the simulator by Aspen Simulation Workbook. The developed Excel and Visual Basic macro based framework allows the comparison of the measured data and the calculated results.

Results and discussion

The process has two inlet streams (C3 fraction and hydrogen) and two outlet streams (hydrogenated liquid C3 fraction and blowdown). Mass flows of every stream are measured. Based on the law of conservation of mass (there is no accumulation):

$$\mathrm{Inlet\ C3} + \mathrm{H_2} = \mathrm{Blowdown} + \mathrm{Product\ C3} \quad (22)$$

If we check whether the measured data satisfy the balance equation, we find that *Eq. (22)* is not fulfilled, so data reconciliation is necessary.

Steps of the procedure are introduced based on mass flow of C3 product fraction. First time random error is separated from basic process using discrete Fourier transformation. *Fig. 7* shows the measured and the filtered values of product C3 fraction. We defined the random noise as the difference between the measured and filtered values so the standard deviation and covariance matrix of random error ($\mathbf{V_{\tilde{d}}}$) can be calculated.

Since there is no accumulation (i.e., $b = 0$) *Eq. (10)* is simplified to

$$\hat{\mathbf{y}} = \left(\mathbf{I} - \mathbf{V_{\tilde{d}}} \mathbf{A}^T \left(\mathbf{A} \mathbf{V_{\tilde{d}}} \mathbf{A}^T \right)^{-1} \mathbf{A} \right) \tilde{\mathbf{y}}. \quad (23)$$

Since two input and two output streams are in the system, the incidence matrix is $\mathbf{A} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}$. Every needed information are known so the matrix operation can be performed. Real industrial data are analyzed, so dimensionless units are included in the figures. The result of reconciliation is visible in *Fig. 8*. The result shows that the proposed method gives a minor improvement of the process values. Thanks to the validated accuracy of the mass measurements, these process values can be directly used for the validation of the simulator.

In the next step, we compare the reconciled output with the values calculated by the simulator in two different ways. First, the input of the simulator is the measured data, then, the reconciled input data. The result of the comparison is shown in *Fig. 9*. First, we thought that the three curves coincide. However, as a portion is zoomed, the difference became apparent (see *Fig. 10*). If the input of simulator is the reconciled data, the calculated and the reconciled output are nearly the same. Although the difference between the two calculated curves may seem small in dimensionless space, do not forget that the small percentage difference can mean significant mass flow values in a real industrial process.
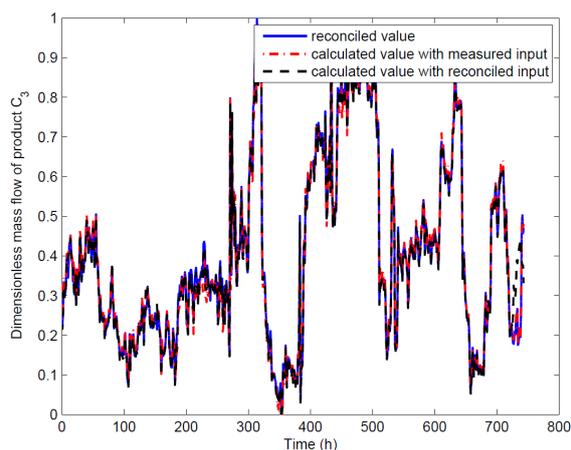
Figure 9: Calculated and reconciled data of product C3 fraction if the input of the simulator is the measured and reconciled data
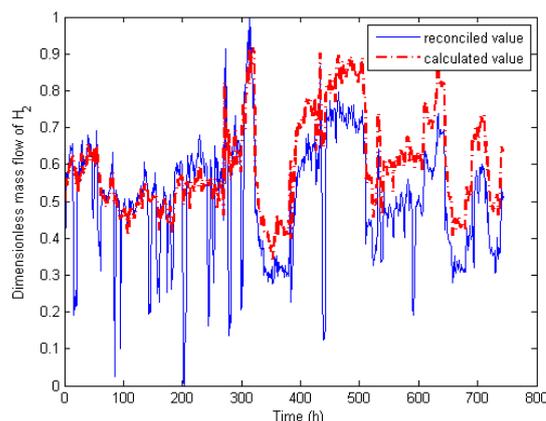
Figure 11: Calculated and reconciled dimensionless mass flow of H_2 if the input of the simulator is the measured data
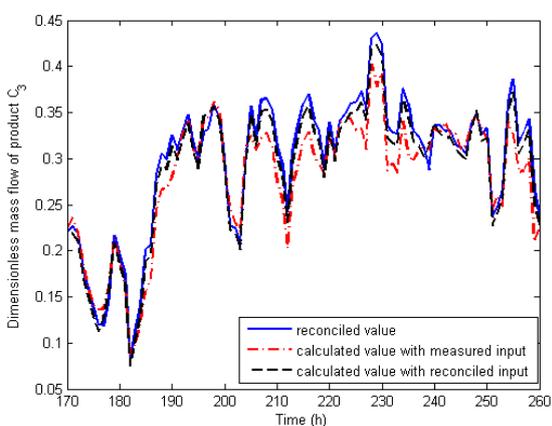
Figure 10: Reconciled and calculated outputs of simulator in case of product C3 fraction

We calculated the square error in both cases (deviation from the reconciled output) and found that the difference decreased by 87% if the reconciled values are the input of the simulator. This shows that the simulator gives more accurate results if previously data are reconciled.

The simulator supported by data reconciliation can be applied in fault diagnosis. *Fig. 11* shows how the hydrogen flow changes over time. The continuous line represents the simulator results without data reconciliation, while the dashed line shows the reconciled values.

Fig. 11 shows the similarity of the calculated and the measured values. After 300 hours of operation the calculated values exceed the measured data. There are several possible reasons. Less hydrogen goes to this system if

- The MAPD content of the inlet stream is lower, because the mass flow of hydrogen is controlled in proportion of the MAPD mole flow of the input C3

fraction.
- After the regeneration when catalyst is active.
- Other reactor is regenerated or reactivated.

In this case, another reactor in the Olefin 2 plant was reactivated (after 300 hours). The reactivation process needed a lot of hydrogen so the hydrogen for the studied reactor was reduced. The received data did not include the rate of reduction. The result of the simulation showed that something was changed in the technology. This shows the technique can be used for monitoring and diagnostics of complex processes because it can be discovered if the operating condition has changed.

Conclusion

Data reconciliation is an effective method to obtain accurate process data that satisfy balance equations using process constraints. The application of steady-state data reconciliation is widespread in the process industry. Simulators are often applied to improve industrial processes, optimize operation, and identify bottlenecks of technology. Historical process data can be used for the identification and verification of models utilized by these tools. Usually, measured data do not satisfy balance equations, because all measurements are incorrect to some extent. Thus, it is necessary to develop a method which can simultaneously and iteratively improve data and model performance. A method based on data reconciliation technique has been developed for this purpose.

An industrial C3 selective hydrogenation process has been analysed as a case study. The steady state simulator of the process has been developed in Aspen Plus flowsheeting software. Calculated and historical process data were compared using Aspen Simulation Workbook.

The proposed technique can be used for monitoring complex processes. Results showed that the proposed technique slightly improves the quality of flow data. Thanks to the accuracy of flow measurements, these process values are directly applicable for the validation of simulator and suitable for further studies, e.g., for the determination of kinetic parameters.

Acknowledgements

SYMBOLS

$\mathbf{A}$	incidence matrix
$\mathbf{A}_y$	incidence matrix of balances in terms of measured flows
$\mathbf{A}_z$	incidence matrix of balances in terms of unmeasured flows
$\mathbf{B}_y$	Jacobian matrix of measured variables
$\mathbf{B}_z$	Jacobian matrix of unmeasured variables
$\mathbf{b}$	source of extensive quantity
$\mathbf{d}$	$\mathbf{y} - \check{\mathbf{y}}$ error
$\mathbf{f}$	function vector of equality model constraints
$\tilde{\mathbf{f}}$	balance error ($\mathbf{A}\tilde{\mathbf{y}} - \mathbf{b}$ vector)
$\hat{\mathbf{f}}$	balance error ($\mathbf{A}\hat{\tilde{y}} - \mathbf{b}$ vector)
$\mathbf{g}$	function vector of inequality model constraints
$\mathbf{I}$	identity matrix
MA	methyl-acetylene
MAPD	methyl-acetylene and propadiene
PD	propadiene
$\mathbf{P}$	projection matrix
$\mathbf{V}_{\tilde{\mathbf{d}}}$	covariance matrix of measurements
$\check{\mathbf{y}}$	vectors of real values of measured variables
$\tilde{\mathbf{y}}$	vectors of measured variables
$\hat{\mathbf{y}}$	vectors of reconciled measured variables
$\check{\mathbf{z}}$	vectors of real values of unmeasured variables
$\tilde{\mathbf{z}}$	vectors of unmeasured variables
$\hat{\mathbf{z}}$	vectors of estimated unmeasured variables
λ	Lagrange multiplier
ΔH_r	reaction heat

REFERENCES

[1] BAUER, M., CRAIG, I. K. Economic assessment of advanced process control - A survey and framework, Journal of Process Control, *2008*, 18(1), 2–18

[2] PACH, F. P. , FEIL, B. , NÉMETH, S., ÁRVA, P. , ABONYI, J. Process data warehousing based operator support system for complex production technologies, IEEE Transactions on Systems, Man and Cybernetics, Part A, Special Issue on 'Advances in Heterogeneous and Complex System Integration', *2006*, 136–153

[3] TIETZE, U., SCHENK, C. Analóg és digitális áramkörök (Műszaki Könyvkiadó)

[4] KUHEN, D., DAVIDSON, H. Computer Control II. Mathematics of Control, Chemical Engineering Progress, *1961*, 57, 44–47

[5] NARASIMHAN, S. Introduction to data reconciliation and gross error diagnosis (2013), URL: www.che.iitm.ac.in/ naras/ ch544/introDRGED.pdf

[6] BELLEC, S., JIANG, T., KERR, B., DIAMOND, M., STUART, P. On-line processing and steady state reconciliation of pulp and paper mill process data, 91st Annual Meeting Preprints-Book B, *2005*, 105–110

[7] THIBAULT, J. Introduction to Data Reconciliation *2013*, URL: www.polymtl.ca/namp/docweb/ Modules_Web/M11_Tier1_Chap1-3.pdf

[8] NOGITA, S. Statistical Test and Adjustment of Process Data, Ind. Eng. Chem, Process Des. Dev., *1972*, 11, 197–200

[9] MAH, R. S. H., STANLEY, G. M., DOWNING, D. M. Reconciliation and Rectification of Process Flow and Inventory Data, Ind. Eng. Chem, Process Des. Dev., *1976*, 15, 175–183

[10] MAH, R. S. H. Chemical process structures and information flows, AIChE Journal, *1993*, 39(4), 730

[11] CAO, S., RHINEHART, R. An efficient method for on-line identification of steady state, Journal of Process Control, *1995*, 5(6), 363–374

[12] TJOA, I. B., BIEGLER, L. T. Simultaneous solution and optimization strategies for parameter estimation of differential − algebraic equation systems, Industrial and Engineering Chemistry Research, *1991*, 30, 376–385

[13] BRITT, H. I. , LUECKE, R. H. The estimation of parameters in nonlinear implicit models, Technometrics, *1973*, 15, 233–247

[14] MADRON, F., VEVERKA, V. Optimal selection of measuring points in complex plants by linear models, American Institute of Chemical Engineering Journal, *1992*, 38(2), 227

[15] CROWE, C. M. Reconciliation of process flowrates by matrix projection. Part II: The nonlinear case, AIChE Journal, *1986*, 32, 616–623

[16] BAGAJEWICZ, M. J., JIANG, Q. Integral approach to plant linear dynamic reconciliation, AIChE Journal, *1997*, 43, 2546–2558

[17] PRATA, D. M., LIMA, E. L., PINTO, J. C. Simultaneous data reconciliation and parameter estimation in bulk polypropylene polymerizations in real time, Macromolecular Symposia, *2006*, 243, 91–103

[18] PRATA, D. M., SCHWAAB, M., LIMA, E. L., PINTO, J. C. Nonlinear dynamic data reconciliation and parameter estimation through particle swarm optimization: Application for an industrial polypropylene reactor, Chemical Engineering Science, *2009*, 64 18, 3953–3967

[19] MARTÍNEZ-MARADIAGA, M., BRUNO, J. C., CORONAS, A. Steady-state data reconciliation for absorption refrigeration systems, Applied Thermal Engineering, *2013*, 51(1-2), 1170–1180

[20] SARABIA, D., DE PRADA, C., GOMEZ, E., GUTIERREZ, G., CRISTEA, S.,SOLA, J., GONZALEZ, R. Data reconciliation and optimal management of hydrogen networks in a petrol refinery, Control Engineering Practice, *2012*, 20(4), 343–354

[21] CHEN, J., PENG, Y., MUNOZ, J. C. Correntropy estimator for data reconciliation, Chemical Engineering Science, *2013*, 104, 1019–1027

[22] KELLER, T., HOLTBRUEGGE, J., GÓRAK, A. Transesterification of dimethyl carbonate with ethanol in a pilot-scale reactive distillation column, Chemical Engineering Journal, *2012*, 180, 309–322

[23] DOMINGUES, L:, PINHEIRO, C. C. , OLIVEIRA, N. M., VILELAS, A., FERNANDES, J., RIBEIRO, F. R. Estimation of catalyst deactivation parameters of ethyl tert-butyl ether (ETBE) reactors based on industrial plant data, in 22nd European Symposium on Computer Aided Process Engineering, *Computer Aided Chemical Engineering*, vol. 30 (I. D. L. BOGLE, M. FAIRWEATHER, eds.) (Elsevier) *2012*, pp. 1002–1006

[24] PRATA, D. M. , SCHWAAB, M., LIMA, E. L., PINTO, J. C. Simultaneous robust data reconciliation and gross error detection through particle swarm optimization for an industrial polypropylene reactor, Chemical Engineering Science, *2010*, 65(17), 4943–4954

[25] VASEBI, A., POULIN, E., HODOUIN, D. Dynamic data reconciliation in mineral and metallurgical plants, Annual Reviews in Control, *2012*, 36(2), 235–243

[26] VALDETARO, E. D., SCHIRRU, R. Simultaneous Model Selection, Robust Data Reconciliation and Outlier Detection with Swarm Intelligence in a Thermal Reactor Power calculation, Annals of Nuclear Energy, *2011*, 38(9), 1820–1832

[27] MARTINS, M. A., AMARO, C. A., SOUZA, L. S., KALID, R. A. , KIPERSTOK, A. New objective function for data reconciliation in water balance from industrial processes, Journal of Cleaner Production, *2010*, 18(12), 1184–1189

[28] STANLEY, G., MAH, R. Observability and redundancy in process data estimation, Chemical Engineering Science, *1981*, 36(2), 259–272

[29] CROWE, C. M. Observability and redundancy of process data for steady state reconciliation, Chemical Engineering Science, *1989*, 44(12), 2909–2917

[30] ALI, Y., NARASIMHAN, S. Redundant sensor network design for linear processes, American Institute of Chemical Engineering Journal, *1995*, 41(10), 2237

[31] ALMASY, G.: Mérlegegyenletek - mérési hibák (Egyetemi Kiadó, Veszprém) *1996*

[32] NAKAYA, M., LI, X. On-line tracking simulator with a hybrid of physical and Just-In-Time models, Journal of Process Control, *2013*, 23(2), 171–178

[33] THIBAULT, J. tate Data reconciliation for Bilinear Systems, *2013*, URL: www.polymtl.ca/namp/docweb/ Modules_Web/M11_Tier1_Chap4-5.pdf

[34] FLORES, A.: Introduction to Aspen Plus (12. 09. 2001)

[35] KANOA, M., NAKAGAWA, Y. Data-based process monitoring, process control and quality improvement: Recent developments and applications in steel industry, Computers and Chemical Engineering, *2008*, 32, 12–24

[36] KADLEC, P., GABRYS, B., STRANDT, S. Data-driven Soft Sensors in the process industry, Computers and Chemical Engineering, *2009*, 33, 795–814

Advertise upcoming meetings,

conferences, and workshops;

make public announcements;

introduce your research laboratory;

a new product or a service

in the

Hungarian Journal of Industry and Chemistry

Please contact us if interested!

STATISTICAL PROCESS CONTROL BASED PERFORMANCE EVALUATION OF ON-LINE ANALYSERS

Tibor Kulcsár[1] and János Abonyi[⊠1]

[1] Department of Process Engineering, University of Pannonia, Egyetem Street 10., H-8200 Veszprém, HUNGARY
⊠E-mail: janos@abonyilab.com

On-line analyzers can provide accurate and timely information for process control and monitoring. Statistical Process Control (SPC) techniques can be effectively utilized to support the development and maintenance of these tools. The D6299-10 ASTM standard details how on-line analyzers should be validated. The applicability of this standard is demonstrated through the analysis of industrial data collected from an on-line gas chromatograph. The results confirm that automatized SPC can effectively improve the reliability of advanced process control systems.

Keywords: on-line analyzer, software sensor, statistical process control, SPC, process monitoring

Introduction

Process variables characterizing and influencing product quality have a significant role in process control and optimization. Off-line laboratory tests mostly take more than two hours. This time delay can cause control problems resulting in economic loss. In such situations, an improved on-line monitoring system is required. On-line analyzers eliminate the dependence on laboratory data. Analysers are valuable instruments for real time control because of their fast response time (1-4 minutes) (see *Fig. 1*) [1].

Quality control techniques can be effectively used to support the development of on-line analyzers [2] and advanced process control systems [3]. The D 6299 ASTM standard (Applying Statistical Quality Assurance Techniques to Evaluate Analytical Measurement System Performance) provides information for the design and operation statistical quality control (QC) tools to monitor and control of analytical measurement systems using a collection of statistical quality control (SQC) tools [4].

The goal of the performance monitoring is the periodic comparison of the on-line analyzer's results to the reference value of the same sample measured by laboratory test methods. Precision and bias (see *Fig. 2*) are calculated to provide information for updating test methods as well as for indicating areas of potential improvements.

Control charts and other statistical techniques can be used for performance monitoring. Statistical estimates of the measurement system precision and bias can be calculated on the basis of periodically updated data. Plotting and interpreting these test results can ascertain the in-statistical-control status of the measurement system [5]. On-line Statistical Process Control (SPC) based real-time validation of measurement systems has been already re-
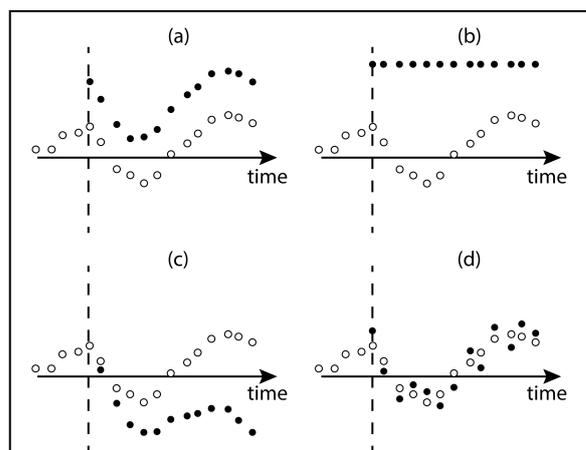
Figure 2: Type of faults. The dashed line shows when the fault occurs. ○: data free of fault; ● corrupted data for the following cases: (a) bias, (b) complete failure, (c) drifting, and (d) precision degradation [4]

Figure 1: Soft sensors and on-line analyzers enable feedback control
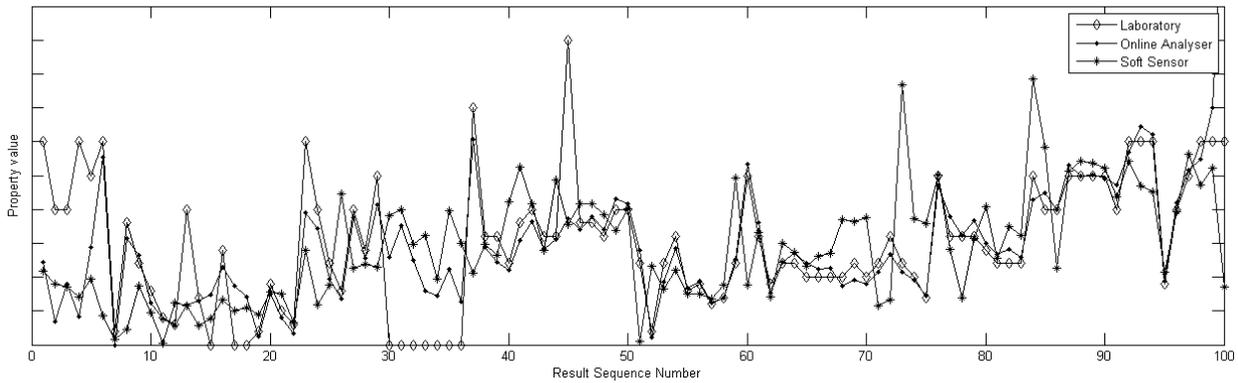
Figure 3: Illustrative example: MOL aromatic block gas chromatograph IP benzene content (m/m %), Lab, Analyser and APC soft sensor

ported in 1997 [6]. The SPC based approach can also be applied for off-line performance evaluation [7]. This approach has a wide range of application areas like steel industry [8], electronic device assembling [9], and buildings' energy demand monitoring [10].

The goal of this paper is to present the theoretical background and application details of SPC based performance evaluation of on-line analyzers. The applicability of the concepts is illustrated by a case study based on data collected from an on-line gas chromatograph of the MOL Plc (see *Fig. 3*) . Typical patters that show out of control status of the system are also presented. To check the normality of the residuals, an easily applicable and interpretable tool is proposed. The developed demonstration tools are available at the website of the authors (www.abonyilab.com).

Control Chart based Evaluation of System Performance

The studied D6299 practice is devoted to a special testing of analyzers [4]. Quality Control (QC) test specimen samples from a specific lot are introduced and tested in the analytical measurement system on a regular basis to establish system performance history in terms of both stability and precision.

The control chart is one of the seven basic tools of QC. Control charts - also known as SHEWHART charts - are used to determine if process is in a state of statistical control. The analysis of the control chart indicates whether the process is currently under control. This means that these charts are used to check the stability of the production. In stable operation, the variations of the process and quality variables are only random, normally distributed variables. In these cases, no corrections of the control parameters are needed. When the chart indicates that the monitored process is out-of-control, the analysis of the chart can help to determine the sources of the variation. Typically, control charts are used for time-series

data, though they can be used for data that have logical comparability [11]. In this section, we present these charts and detail how these should be applied for the performance assessment of on-line analyzers.

Control Charts

Run Chart

The run chart is a plot of sample values in chronological order (*Fig. 4*). The run chart can be used to screen data for unusual patterns such as continuous trending in either direction, unusual clustering, and cycles. The run chart of the data used in our case study is shown in *Fig. 3*. The plotted time series shows the signal of the on-line analyzer and the related laboratory measurements.

I Chart

The I (individual) chart is a run chart to which control limits and center line have been added (see top panel of *Fig. 5*). The center line is based on the mean of the samples,

$$\bar{I} = \frac{\sum_{i=1}^{n} I_i}{n} \tag{1}$$
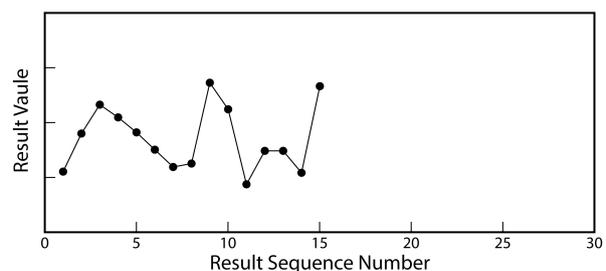
Figure 4: Example for a run chart
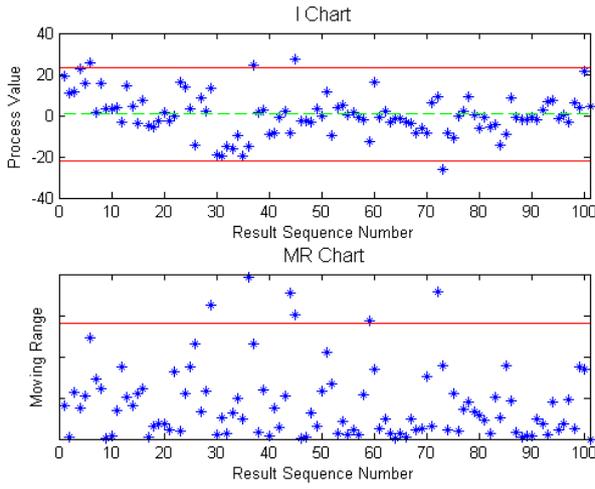
Figure 5: I and MR charts show out of control samples of the gas chromatograph IP benzene content (m/m) %

Figure 6: Example for a MR chart
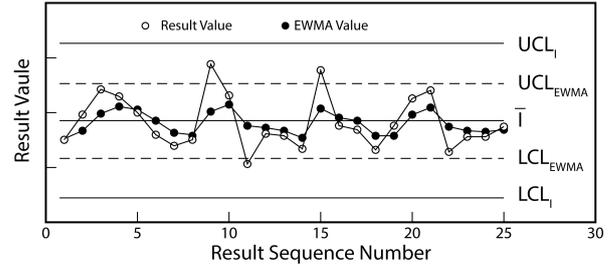
Figure 7: Example for an EWMA chart

while the upper and lower control limits are based on the estimated variance (range of the data) $\overline{MR}$ [12]

$$\overline{MR} = \frac{\sum_{i=1}^{n-1} |I_{i+1} - I_i|}{n-1}, \qquad (2)$$

$$UCL_I = \bar{I} + 2.66\overline{MR}, \qquad (3)$$

$$LCL_I = \bar{I} - 2.66\overline{MR}. \qquad (4)$$

Individual values that are outside the upper or lower control limits are indications of an unstable system, and efforts should be made to determine the cause [5]. Optionally, any one of the following occurrences should be considered as potential signs of instability:

1. Two out of three consecutive results on the I chart that are more than $1.77\overline{MR}$ distant from the center line in the same direction;

2. Five consecutive results on the I chart that are more than $0.89\overline{MR}$ distant from the center line in the same direction;

3. Eight or more consecutive points in the I chart that fall on the same side of the center line.

MR Chart

MR (Moving Range) charts are also used to detect unusual patterns by plotting the sequential range of two values given by

$$\Delta_i = MR_i = |I_i - I_{i-1}|, \qquad (5)$$

and connecting each point (see *Fig. 6* for an example). There is no lower control limit for an MR chart [5]. The upper control limit for the MR chart is given by

$$UCL_{MR} = 3.27\overline{MR}. \qquad (6)$$

EWMA chart

An EWMA (Exponentially Weighted Moving Average) chart is used to enhance the sensitivity in detecting mean shifts that are small relative to the measurement system precision (see *Fig. 7* for an example). Each $EWMA$ value is a weighted average of the current result and previous results with the weights decreasing exponentially with the age of the reading [5]:

$$EWMA_1 = I_1, \qquad (7)$$

$$EWMA_i = (1 - \lambda)EWMA_{i-1} + \lambda I_i, \qquad (8)$$

where λ is the exponential weighting factor. For application of this practice, a λ value of 0.4 is recommended.

The control limits for the EWMA chart are calculated using a weight (λ) as follows:

$$UCL_\lambda = \bar{I} + 2.66\overline{MR}\sqrt{\frac{\lambda}{2-\lambda}}, \qquad (9)$$

$$LCL_\lambda = \bar{I} - 2.66\overline{MR}\sqrt{\frac{\lambda}{2-\lambda}}. \qquad (10)$$

The complexity of multivariate and autocorrelated processes makes it difficult to use standard control charts. To construct simple and interpretable charts, dimensional reduction could also be used as SIMOGLU and MARTIN have done [13], but in case of autocorrelated data, model-based control charts should be applied as KIM and JITPI-TAKLER have done in their research [14].

Pretreatment of Test Results

Assessment, control charting, and evaluation are applied only to appropriately pretreated test results. The purpose

of pretreatment is to standardize the control chart scales so as to allow for data from multiple check standards to be compared on the same chart. For QC sample test results, no data pretreatment is typically used since results for different QC samples are generally not plotted on the same chart.

In our case, the difference between the measurements and their accepted reference values (ARVs) are monitored. ARV serves as an agreed-upon reference for comparison and that is derived based on (1) a theoretical value, based on scientific principles, (2) an assigned value, based on experimental work, or (3) a consensus value, based on collaborative experimental work under the auspices of a scientific or engineering group.

$$I = \text{test result} - ARV. \tag{11}$$

Assessment of Initial Results

In the initial phase of the application assessment techniques are applied to test results collected during the startup phase of or after significant modifications to a measurement system. It is required to perform the following assessment after at least 15 pretreated results have become available. The purpose of this assessment is to ensure that these results are suitable for deployment of control charts.

Pretreated results should first be visually screened for values that are inconsistent with the remainder of the data set, such as those that could have been caused by transcription errors. Those flagged as suspicious should be investigated. Discarding data at this stage must be supported by evidence gathered from the investigation. If after discarding suspicious pretreated results there are less than 15 values remaining, collect additional data and start over.

The next step is to examine the pretreated results for non-random patterns such as continuous trending in either direction, unusual clustering, and cycles. One way to do this is to plot the results on a run chart and examine the plot. If any non-random pattern is detected, investigate for and eliminate the root cause(s).

Typical Control Chart Patterns

In the previous sessions, the main charts used for performance measurement were presented. In SPC, the Western Electric Rules are the decision rules for detecting "out-of-control" or non-random conditions on control charts. Locations of the observations relative to the control chart control limits (typically at ± 3 standard deviations) and centerline indicate whether the process in question should be investigated for assignable causes. The Western Electric Rules were codified by a specially-appointed committee of the manufacturing division of the Western Electric Company and appeared in the first edition of its Statistical Quality Control Handbook in 1956 [15]. Their
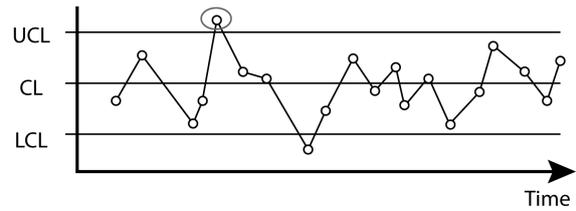
Figure 8: This process is out of control because a point is either above the UCL or below the LCL. For example, in Fig. 5-B at sequence number 28 there is a unique point below the LCL
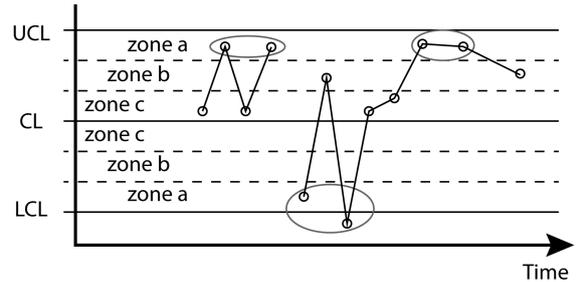
Figure 9: In this case the system produce 2 out of 3 consecutive points either in or beyond zone A. The process on Fig. 5-B shows this behaviour at sequence number 45 where the values are close to the Upper Control Limit.

purpose was to ensure that line workers and engineers interpret control charts in a uniform way [5, 12]. The eight standard Western electric rules are:

1. The most recent point plots outside one of the 3-sigma control limits (see Fig. 8). If a point lies outside either of these limits, there is only a 0.3% chance that this was caused by the normal process.

2. Two of the three most recent points plot outside and on the same side as one of the 2-sigma control limits (see Fig. 9). The probability that any point will fall outside the warning limit is only 5%. The chances that two out of three points in a row fall outside the warning limit is only about 1%.

3. Four of the five most recent points plot outside and on the same side as one of the 1-sigma control limits. In normal processing, 68% of points fall within one sigma of the mean, and 32% fall outside it. The probability that 4 of 5 points fall outside of one sigma is only about 3%.

4. Eight out of the last eight points plot on the same side of the center line, or target value. Sometimes you see this as 9 out of 9, or 7 out of 7. There is an equal chance that any given point will fall above or below the mean. The chances that a point falls on the same side of the mean as the one before it is one in

two. The odds that the next point will also fall on the same side of the mean is one in four. The probability of getting eight points on the same side of the mean is only around 1%.

5. Six points in a row increasing or decreasing. The same logic is used here as for rule 4 above. Sometimes this rule is changed to seven points rising or falling.

6. Fifteen points in a row within one sigma. In normal operation, 68% of points will fall within one sigma of the mean. The probability that 15 points in a row will do so, is less than 1%.

7. Fourteen points in a row alternating direction. The chances that the second point is always higher than (or always lower than) the preceding point, for all seven pairs is only about 1%.

8. Eight points in a row outside one sigma. Since 68% of points lie within one sigma of the mean, the probability that eight points in a row fall outside of the one-sigma line is less than 1% (see *Fig. 10*).

Normality Checks

Since the control chart and limits prescribed in this practice are based on the assumption that the data behavior is adequately modeled by the normal distribution, it is recommended that a test of this normality assumption be conducted. One way to do this is to use a normal probability plot and the ANDERSON-DARLING Statistic [16]. If the results show obvious deviation from normality, the statistical control charting techniques described are not directly applicable to the measurement system [5].

Quantile-quantile plot (q-q plot) is a graphical tool for comparing two probability distributions by plotting their quantiles against each other. The normality plot of the process can be obtained by comparing the empirical distribution of the data against a standard normal distribution

Figure 11: Normality check of the studied gas chromatograph.

(see *Fig. 11*). When the results are normally distributed, the plot should be approximately linear. Major deviations from linearity are an indication of non-normal distributions of the differences [17, 16].

Conclusions

We showed that statistical process control can be effectively used to support the development and maintenance of on-line process analyzers. A case study is presented based on the analysis of data taken from the chemical process industry. The proposed concept has been implemented in MATLAB. The results illustrate the applicability the developed tools in improving the reliability of advanced process control systems.

Acknowledgements

Figure 10: Long runs (8 or more consecutive points) either above or below the centerline. (See *Fig. 5*-A in range 28-40)

REFERENCES

[1] GIDWANI, K. K. , BECKMAN, R. F. Evaluation of refinery control systems, ISA Transactions, *1994*, 33(3), 217–225

[2] ROBERTO, M. F., DEARING, T. I. , MARTIN, S., MARQUARDT, B. J. Integration of Continuous Flow Reactors and Online Raman Spectroscopy for Process Optimization, Journal of Pharmaceutical Innovation, *2012*, 7(2), 69–75

[3] SMITH, C. Advanced Process Control: Beyond Single Loop Control (John Wiley & Sons), *2011*

[4] A. INTERNATIONAL: Applying Statistical Quality Assurance Techniques to Evaluate Analytical Measurement System Performance (ASTM International), *2010*

[5] MONTGOMERY, D. Introduction to statistical quality control (John Wiley & Sons, Inc.), *2009*

[6] CONVERSE, J. Real time validation of on-line measurements utilizing internal standard introduction, ISA transactions, *1997*, 36(2), 151–158

[7] COLLEDANI, M., TOLIO, T. Performance evaluation of production systems monitored by statistical process control and off-line inspections, International Journal of Production Economics, *2009*, 120(2), 348–367

[8] KANO, M., NAKAGAWA, Y. Data-based process monitoring, process control, and quality improvement: Recent developments and applications in steel industry, Computers & Chemical Engineering, *2008*, 32(1-2), 12–24

[9] REIS, M. S., DELGADO, P. A large-scale statistical process control approach for the monitoring of electronic devices assemblage, Computers & Chemical Engineering, *2012*, 39, 163–169

[10] BRAGA, L., BRAGA, A. R., BRAGA, C. On the characterization and monitoring of building energy demand using statistical process control methodologies, Energy and Buildings, *2013*, 65, 205–219

[11] POOTS, A. J. , WOODCOCK, T. Statistical process control for data without inherent order, BMC Medical Informatics & Decision Making, 12, *2012*

[12] CHANDRA, M. Statistical quality control (CRC Press LLC), *2001*

[13] SIMOGLOU, A., MARTIN, E., MORRIS, A. Multivariate statistical process control of an industrial fluidised-bed reactor, Control Engineering Practice, *2000*, 8, 893–909

[14] KIM, S. B. , JITPITAKLERT, W., PARK, S.-K. , HWANG, S.-J. Data mining model-based control charts for multivariate and autocorrelated processes, Expert Systems with Applications, *2012*, 39(2), 2073–2081

[15] THOMAS, D., ET AL. Statistical Quality Control Handbook (Western Electric Co., Inc), *2011*

[16] THODE, H. C. Testing For Normality (CRC Press LLC), *2002*

[17] HARRIS, T., SEPPALA, C., DESBOROUGH, L. A review of performance monitoring and assessment techniques for univariate and multivariate control systems, Journal of Process Control, *1999*, 9(1), 1–17

CORROSION PROTECTION WITH ULTRATHIN GRAPHENE COATINGS: A REVIEW

ANDRÁS GERGELY ✉ AND TAMÁS KRISTÓF

Department of Physical Chemistry, Institute of Chemistry, University of Pannonia, Egyetem u. 10.,
Veszprém 8200, HUNGARY
✉Email: gergelyandras@almos.uni-pannon.hu

Developments in surface treatment or finishing, and modification of structural metals have probably never been so dynamic than in recent years due to the need for new approaches to efficient corrosion protection. The application of coatings is a strategy to be followed to physically separate corrosive environments from metal surfaces. The overall protection efficiency depends on the coating's barrier properties. Traditional alternatives to coating suffer from inefficient physical protection in cases of low film thicknesses and at elevated temperatures. One of the most advanced options is to apply ultrathin atomic films to ensure complete separation of the metallic surface from the fluid media. Among the numerous materials and methods, exceptional chemical resistance and high domain size make graphene a promising candidate for constituting ultrathin coatings with interfacial atomic layers adherent and homogeneous coverage to feature firm barrier behaviour. This review focuses on the major efforts with notable results and points out some short comings that must be resolved to serve as a basis for further progress in this field.

Keywords: ultrathin coatings, graphene, metals, corrosion protection

Introduction

Nowadays graphene (Gr) is undoubtedly of great interest in many fields of science and engineering. The attention is due to its unique structure, electrical properties [1, 2], efficient heat spreading [3], wetting transparency [4], surface wetting tenability [5], and high reflectivity (mirrors for the beams of He and H_2) [6]. The Gr's practical impermeability to gases [7-9] is a result of the combination of its strong structure (carbon bond energy and intrinsic strength are 4.9 eV and 43 N m^{-1}, respectively) and its morphology that makes Gr to be the thinnest and most impermeable membrane [10]. In addition, because of its chemical inertness [11], especially against oxidation [12, 13], it is highly resistant to corrosion [14] under conditions where other substances would undergo chemical reactions [15, 16]. As a lightweight mono-atomic coating, Gr can function as one of the most advanced electron- and ion-transfer barriers at the metal-electrolyte interface, without notably altering optical properties and thermal conductivity of the substrates. Gr on metal surfaces is expected to afford oxidation resistance [11] based on its airtight balloon behaviour associated with its excellent barrier nature against oxygen diffusion [7]. Nevertheless, coherence, and uniformity of Gr coatings are apparently vital to achieve good corrosion protection. In accordance with general expectations, electrochemical properties of Gr [17, 18] and its coated derivatives were extensively investigated. The main results achieved recently are described in the following sections that are organised from the viewpoint of synthesis techniques and application purposes.

Pure Graphene Coatings

Direct Chemical Vapour Deposition (CVD) on Metal Surfaces

Early studies were inspired to synthesise full surface covering Gr films of the highest domain (single sheet) size and integrity to exert impermeability over macroscopically large surfaces. Later findings proved the viability of this concept as interfacial construction of thin layers is a promising way to provide durable corrosion protection for metallic substrates [19]. Generally, coatings play a vital role in improving surface quality and protection of substrates, since much effort has been made to produce films featuring excellent properties. Graphene is certainly one of the most prominent representatives of this kind and subsequently has been studied in almost all scientific and engineering fields. This new generation material has been a novel subject to incorporate into the latest corrosion protecting coatings. Achievements in utilising Gr as a pure and composite coating were first reviewed by TONG et al. [20]. The synthesis and functionalisation of Gr were briefly discussed and future perspectives suggested.

The simplest way is to grow Gr directly on metal surfaces, even though this strategy does not provide the most protective coatings due to low adherence of the inorganic layers as consequence of weak interaction. By direct CVD, metals like nickel [21], cobalt [22], iridium [23], ruthenium [24] and copper [21, 25-27] were coated with Gr (on large areas in certain cases) [25] using methane or acetylene as a carbon source [28, 29]. Deposition kinetics of carbon and the formation of Gr layers on copper were thoroughly explored; surface mediation theory was proved when poly-crystalline films were characterised with the domain size of 10 μm^2 [30]. Except for assessing the influence of grain orientation, the effect of other growth parameters was estimated. The most relevant findings are as follows:

1. The density of Gr nuclei decreased, conversely flake domain size increased as temperature increased and partial pressure of the carbon source decreased.
2. When the pressure dropped below a critical threshold then no Gr nucleation was observed and film formations terminated before reaching full surface coverage.
3. To achieve complete surface coverage, the partial pressure of methane must have been increased above 500 mTorr then full coverage was reached within 2 to 3 min. Only a supersaturated surface relaying on a critical threshold of partial pressure ~285 mTorr was suggested to drive the formation of fully coated copper substrates with a continuous layer of Gr flakes. Thus, carbon nuclei formed in the first step under a lower pressure followed by a second step of nuclei growth at a higher pressure, promoting full surface coverage.

The effects of these parameters coupled with the investigation of the orientation of metal crystals on Gr domain size were quantitatively evaluated at low temperatures [31]. The activation energy for Gr nucleation under atmospheric pressure was noticeably higher (9 eV) than under low-pressure CVD procedures (4 eV). Such a difference was attributed to copper sublimation at low pressure, because copper evaporation probably has a decisive impact on the desorption rate of carbon from the surface. Copper evaporation is restricted under atmospheric pressure, so the activation energy is assigned to the desorption of carbon clusters. Close to the melting point of copper, large single-crystal Gr synthesis was proposed at the highest possible temperature (probably to maintain the highest rate of interfacial mass-transfer, diffusion coefficient of the decomposed and fragmented precursor agents). By following these principles, single Gr crystals with a domain size of ~1 mm could be synthesised. The nucleation density was higher on (111) than on (100) and (101) surfaces without any identifiable preference at higher temperatures [32]. Under carefully optimised conditions, continuous films of some layers were obtained on polycrystalline nickel without void areas at the flake boundaries [21]. In that case, ambient pressure CVD was to grow Gr films of 1-12 layers, with flake sizes of up to 20 μm for single- and bilayers (BLs). An ultra high vacuum (UHV) chamber was not needed [33, 34] to finely balance nucleation and growth processes,

instead, thermal annealing was employed leading to the emergence of single-crystalline nickel with grain sizes of 1-20 μm. Grain surfaces became atomically flat with some terraces and steps similar to single-crystals usually used for epitaxial UHV experiments and Gr growth took place probably *via* the precipitation of carbon in the metal. Because metallic grains may independently affect film thickness as a result of the high density of atomic steps at grain boundaries, multi-layer (ML) nucleation occurred mainly at those areas.

By using atmospheric pressure CVD, high quality Gr was grown on copper, eliminating all the major difficulties associated with the low-pressure method. Parameter optimisation involved evaluating the influence of the thickness, purity, morphology and crystallographic orientation of the metal foils on growth rates and the number of Gr layers. Scanning electron microscopy (SEM) images illustrating Gr flakes on polished foils besides XRD patterns of annealed foils are presented in *Fig.1* showing large copper domains on NR foils. The effects of copper impurities on the density of bilayers are signified; complete surface coverage on AA2 (the purest foil (99.999%) without double layers), AA1 and NR. In *Fig.2*, Gr quality is presented at various distances along the tube reactor. The methane concentration was higher at the inlet leading to a greater abundance of MLs (darker islands in the images). Gradual increase of concentration resulted in the formation of single-layer (SL) Gr growth over the entire length. A constant methane concentration yielded a high percentage of MLs at the inlet, incomplete coverage in the middle and separated single domains (seen as darker islands) at the outlet.

As a method of surface pre-treatment, electro-polishing [36, 37] in phosphoric acid seemed to be a good choice to achieve an appropriate degree of surface cleaning and minimised roughness. Acetic acid was highly efficient at cleaning the copper surface, but only capable of removing oxidised species without any further etching. Etching with ferric chloride solution generally gave similar results to electropolishing despite the greater roughness of the $FeCl_3$ treated copper foils. This was probably due to the considerable scale of copper recrystallisation during annealing at temperatures close to the melting point of copper before the CVD process, leading to the evolution of a smoother surface with a low density of defects serving as nucleation sites. Additionally, the growth rate was estimated depending on the purity and crystallographic orientation of copper but by far not as much as the partial pressure of hydrogen, because the amount of dissolved hydrogen depends less on the thickness of the copper substrate. Crystallographic orientation was found to be an important factor, since the growth rate on the surface of substrates characterised by Miller indexes (100) was faster than on (111) surface under ambient pressure. In contrast, at low pressure and temperature the (111) surface facilitates the fastest growth rate [38]. Furthermore, optimal conditions to achieve full surface covering films with the minimum proportion of MLs might greatly vary with purity of the metal specimens. High purity substrates contribute to the formation of SL
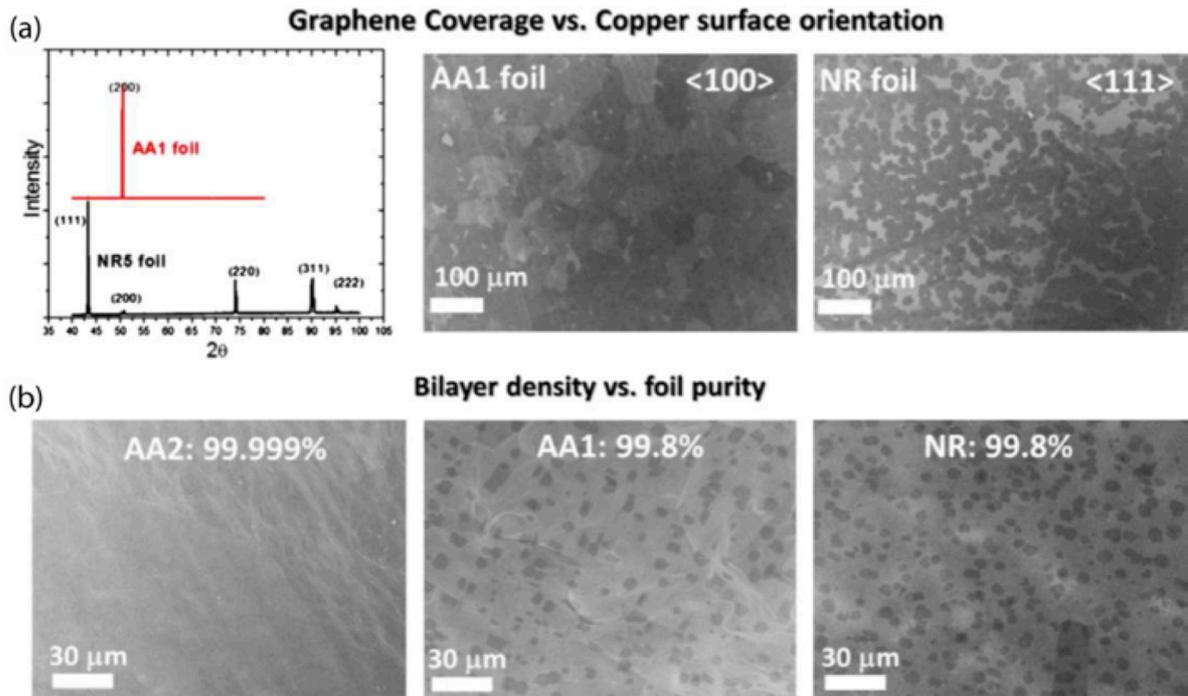
Figure 1: SEM images illustrate (a) Gr growth rate on polished AA1 and polished NR foils. XRD of annealed AA1 shows (100) orientation, the NR foil with the majority of (111) with smaller contributions from (110) and (311); (b) The effect of copper impurities on the density of bilayers [35] (republished by courtesy of the publisher, Elsevier)

films, and low purity metals support the evolution of multi-layer (ML) Gr coatings. Due to the multiple aspects of the reaction mechanism, optimal conditions for long foils are related to low and varied concentrations of carbon sources along the tube reactors, similarly to the application of temperature gradients (furnace dimensions and its relation to foil size) as well.
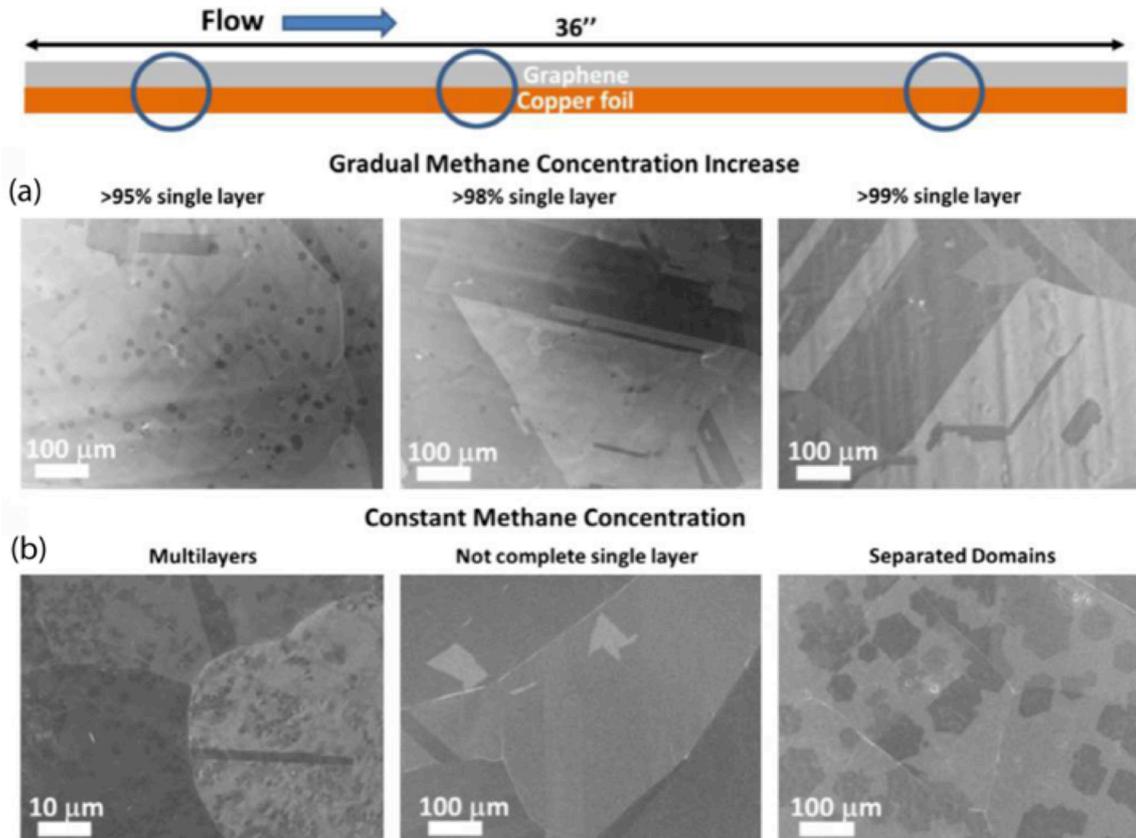
Figure 2: Gr films at various distances along the CVD reactor used at atmospheric pressure. (a) Gradual methane concentration, single-layer Gr growth over the tube reactor; (b) Multilayers at the inlet, incomplete surface coverage in the middle and separated single domains at the outlet [35] (republished by courtesy of the publisher, Elsevier)
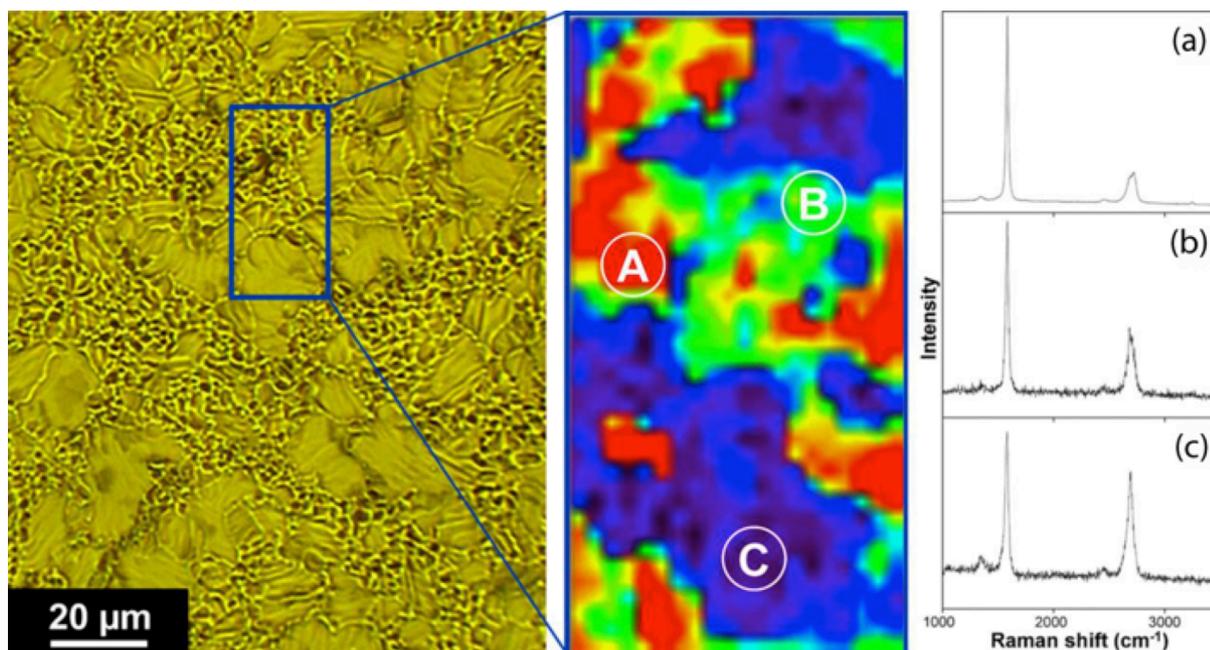
Figure 3: Optical micrograph, Raman map of ~2700 cm^{-1} peaks and respective Raman spectra of the areas on the Ni–Gr surface [41] (republished by courtesy of the publisher, Elsevier)

The gradual increase of methane concentration from 30 to 100 ppm in the mobile phase helps resolve the changing rate of activated carbon production [39]. Because methane concentration near the outlet increases slowly as a result of thermal decomposition and deposition, the main purpose of the gradually increasing methane concentration was to minimise bilayer growth at the inlet over the synthesis duration and to ensure full surface coverage near the outlet with the highest concentration of the carbon source. In fact, optimisation in the case of the atmospheric pressure CVD is more complicated than vacuum ones, which process in long tube chambers with small foils. A steady methane concentration at the inlet surely results in an increasing proportion of MLs at the inlet and incomplete coverage at the outlet, with visibly separated flake boundaries. The initial methane concentration of less than 30 ppm exclusively led to SL Gr at the inlet. High quality Gr grows with >99% monolayer coverage and in some cases films consisted of flake sizes of around 100 μm. High temperature etching with oxygen led to the visualisation of inter-domain boundaries as etching started at these boundaries and expanded into domains in a dendritic pattern, leaving diagonal stripes etch-free. Efficient corrosion resistive film properties were reported when experiments were made with samples exposed to air at 300 ºC, showing no sign of oxidation (oxygen enrichment in the metallic phase) over Gr coated areas in comparison to bare substrates.

According to a similar preference, growing Gr layers on copper is favourable, because copper does not dissolve carbon as much as nickel at high temperatures. However, it is also possible to grow MLs with highly ordered crystalline structures suing atmospheric CVD with controlled film thickness [40]. Both pure nickel and copper were covered with CVD grown Gr films detected as covering the majority of the metals in a SL form besides some areas of double and fragmented MLs

(less than 5% of the coating) [41]. An optical micrograph and Raman map with the respective Raman spectra obtained from the Ni–Gr and Cu–Gr surfaces are presented in *Figs.3* and *4*, respectively. Electrochemical tests indicated that such coatings substantially reduced the corrosion rate of the substrates. By Raman spectroscopy investigation, SL Gr showed sharp G (1580 cm^{-1}) and 2D (2650–2700 cm^{-1}) bands with a small G/2D ratio, whereas ML Gr with a high G/2D ratio and an altered, less sharp 2D band were detected. Hence, 80 and 60% of the copper/nickel surface was covered with Gr as a single- or few-layers coating without uniformity on nickel. Scanning electron micrographs of the samples showed different thicknesses of the films with various scales of folding, full of wrinkles and edges. Thus, the overall conclusion was reached that coherent SL Gr is hard to obtain by direct deposition processes. Immersion type corrosion tests indicated open circuit potentials (OCPs) or free corrosion potentials of copper and nickel shifted towards the positive potential region. The electrochemical test conditions are summarised in *Table 1*. A potential ennoblement of about 0.3 V was measured for the corroding system of nickel/Gr and a cathodic shift of ~0.2 V observed with coated copper samples. The cathodic reduction processes were affected, inhibited by the Gr layer on copper and the rate of anodic half-reactions was apparently reduced near the steady free corrosion potential. However, during extended linear potential scanning Gr/nickel showed higher anodic currents than uncoated substrates opposed to the modified copper that exhibited no changes in the anodic branch after Gr deposition.

As an oxidation barrier for liquid and liquid–vapour phase-change cooling systems almost the same heat transfer efficiency was reported using bare and Gr/copper surfaces in liquid-phase and two-phase thermal performance tests [42]. After thermal treatment,
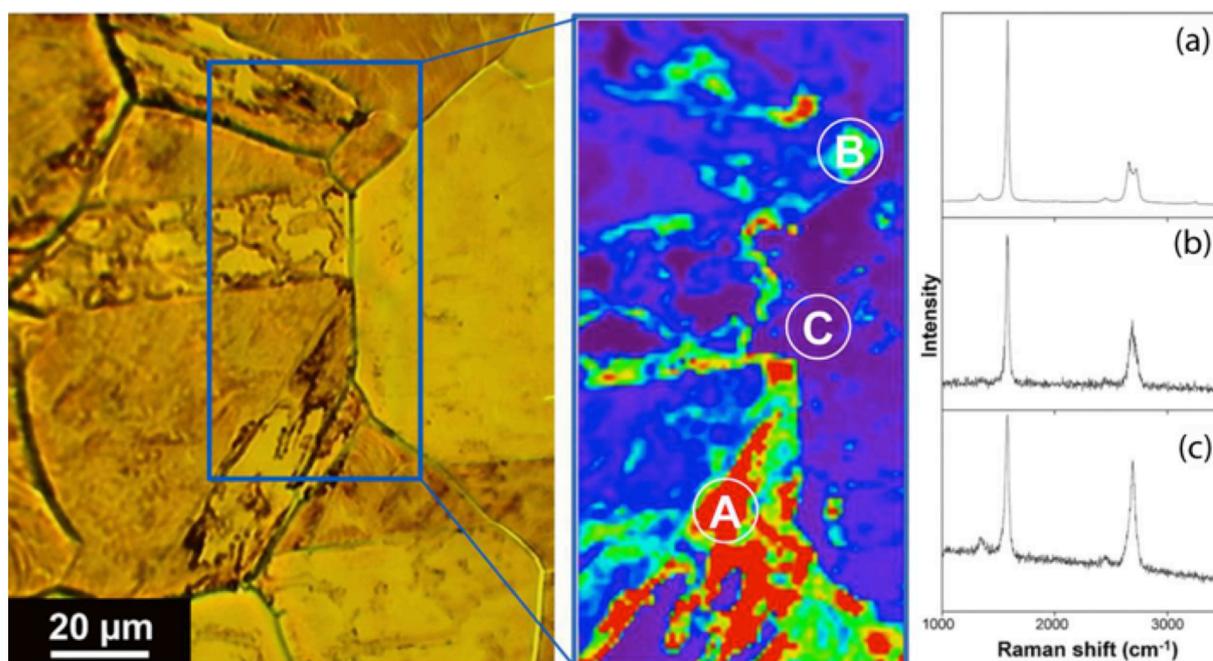
Figure 4: An optical micrograph, Raman map of ~2700 cm^{-1} peaks and respective Raman spectra of the areas on the Cu–Gr surface [41] (republished by courtesy of the publisher, Elsevier)

analysis revealed oxide formation on the entire surface of bare copper and limited oxidation at grain boundaries on the coated substrate. In conclusion, few Gr layers acted as a protective layer even under vigorous flow conditions near the boiling point of the fluid phase. Graphene prepared by microwave (MW) assisted CVD protected substrates from chemical oxidation when samples were heated under atmospheric conditions in a sequential multi-step procedure with dwelling times from 15, or 30 min to 1 h [43]. A very low degree of chemical oxidation of Gr over copper was indicated by XPS and the presence of cuprous oxide was found on the entire surface of bare specimens.

There are special applications such as steering and focusing elements for neutral atomic and molecular beams in de Broglie microscopes [44]. These mirrors utilise the quantum reflection of an atomic beam on metal surfaces since interaction volume is strictly limited to an extremely thin atomic layer of the metals. Oxidation and any other form of corrosion or adsorption of atmospheric gases might well lead to far lower reflectivity of the metallic surfaces, and the coatings of more than a few atomic layers are unacceptable. Thus, there are exceptional requirements of protective coatings and Gr films were expected to provide the solution in this matter too. Complete and uniform monolayers could be grown on polycrystalline ruthenium of a columnar structure on fused silica, exposing flat surface facets [45]. Single-layer Gr protects underlying metals with special structures like concave focusing mirrors and non-planar micro-electrode arrays from oxidation under ambient gases. Ruthenium was recovered by heating the samples under ambient conditions at 250 ºC for half an hour in an ultra-high vacuum, which was indicated by missing peaks in the oxygen (1s) region of the XPS spectra.

Gr on copper provided mediocre protection against electrochemical degradation in chloride environments [46]. Impedance measurements showed increased resistance at the metal/solution interface as anodic and cathodic current densities decreased with about hundred times higher charge-transfer resistance besides about thousand times lower double-layer capacitance in comparison to uncoated copper. Ennoblement of the free corrosion potential was also observed and related to the good barrier nature of the film.

The protection of copper against corrosion by the growing number of Gr layers was in part demonstrated [47]. To enable reliable and complete mechanical blockage by the coating exploiting its impermeability and chemical inertness, preparation parameters were balanced to obtain an appropriate film structure. In *Fig.5*, cyclic voltammograms (CVs) of the bare and passivated copper (the latter with 1-3 Gr layers) are shown along with the intensity of oxidative currents besides the atomic force microscope image of the Cu/Gr after etching.

Fig.6 depicts CVs of Gr-coated copper and Gr-copper post-treated with aluminium oxide (AlO$_x$) of different thicknesses, along with the comparison of their peak currents. Nonetheless, nanometre size structural defects at flake boundaries were the main source of deficiency, decisive in the often-observed limited protection efficacy. To circumvent this problem, a general strategy is to grow ML island-free Gr films. Selective coverage of defect areas was achieved by atomic layer deposition, resulting in enhanced protection of ~99%. Bare copper in an aerated sulphate solution indicated a pitting corrosion rate of ~9 nm s^{-1} with a density of ~1 μm^{-2}.

The estimated diffusion rate of the reactants was above 1 m s^{-1} limiting the rate of corrosion processes.

Table 1: Metallic substrate, type and growth parameters to obtain Gr films by direct CVD deposition

Substrate	Preparation type & parameters	Corrosion test parameters	Surface of the working electrode (cm^2)	Types of electrochemical test	Ref.
Cu & Ni	CVD grown on metal surfaces	flat-cell, 0.1 M NaCl	0.07	potentiodyn. pol., scan rate: 1 mV s^{-1}, after min at OCP	[41]
Cu	CVD grown	flow-cell, mass velocity: 38 kg m^{-2} s^{-1} for 9 h, T = 59.5±0.5°C	3.06	none	[42]
Cu	MW assisted CVD	heating at 150°C, in air at 1 atm for 1 h	Unknown	none	[43]
Ru	CVD	ambient exposure	Unknown	none	[45]
Cu	CVD	0.1 M NaCl	0.785	potentiodyn. pol., scan rate: 0.5 mV s^{-1} from OCP, EIS after 1 h	[46]
Cu	CVD	aerated 0.1 M Na$_2$SO$_4$	0.24	cyclic voltammetry, scan rate: 10 mV s^{-1}	[47]
Ni	CVD	treated at 600 °C in air, then 31 wt.% H$_2$O$_2$ for up to 2 h	Unknown	none	[49]
Cu & Ni	CVD & mech. transferred by spin-coating with PMMA at 4000 rpm for 45 s	0.1 M NaSO$_4$	0.4	potentiodyn. pol. & CV, scan rate: 0.005 mV s^{-1}, EIS	[50]
Cu & Si	CVD then transfer of layers with PMMA cured at 165 °C	ox. at 225 °C for 60 min and red. of Cu at 450 °C for 90 min under O$_2$ and H$_2$ flow	Unknown	none	[60]
Cu	CVD, thiol func.: SLG-Cu treated with HCl & H$_2$O then 7.5 mM thiol-ethanol treatment for 2 h	lymphocyte transformation test & *in vivo* experiment	1.4	tests: Cu conc. by ICP-OES, cell viability & stimulation index for lymphocyte	[80]
Ni	template directed CVD	anode in glucose based electrolyte, Ni in ferricyanide catholyte	8	CV and EIS	[116]

Interestingly, in the case of the ML Gr films, mass transport remained fast across the sheets and parallel along their planes, which was anticipated by the outcome of a water permeation experiment [8]. This may explain the unchanged chemical inertness of the Gr coating, owing to the fast mass transport along and across the layers as etchant species permeate through unburied areas in outer layers and diffuse laterally while finding open areas in layers underneath. Inertness to electrochemical reactions could only be attained by thin layer deposition of aluminium oxide (within ~5 nm) by the means of an atomic layer deposition technique on Gr/copper. Only by this way could unburied regions on the copper surface be made void-free. After 160 atomic layer deposition cycles with a ~16 nm thick passivating film inhibition efficiency was ~99%. This was partly confirmed by Tafel analysis giving instantaneous general corrosion rates of ~99% to the copper-single layer Gr with 16 nm aluminium oxide and 87% to SL Gr/Cu, in comparison with bare copper.

The approach of the direct loading of a precursor, i.e., acetone on copper then annealing it to convert its vapours into Gr was investigated [47]. Monolayers formed with nearly 100% coverage after rapid thermal annealing. Under optimal conditions, acetone derived films with good crystallinity compared to common CVD grown counterparts. The passivity of the metal coated/masked with SL Gr was attributed to remarkably inhibited cathodic reduction reactions and barrier action against the diffusion of anions accessing the underlying substrate. So, SL Gr/copper exhibited ca. forty times lower corrosion rates compared to the mechanically polished copper, which means an inhibition efficacy of up to 97%. Nevertheless, optimal protection performance was proposed to be achieved by the application of four Gr layers with their short-term inhibitions of up to 99%. The corrosion current of the Gr/copper reduced by ca. sixty times compared to mechanically polished substrates and the OCP shifted towards the negative potential region. Electrochemical impedance spectroscopy (EIS) indicated ten thousand times lower electrolyte-metal interfacial capacitance (C_{dl}) of the Gr/copper samples than those of the mechanically polished specimens. Such a notable drop suggests a significant shrinkage of the exposed surface to the electrolyte. C_{dl} of the Gr/copper was ca. three times higher in NaCl solution (0.6 M) than in its diluted media (0.1 M), but 6.4 times lower than that obtained in 0.1 M Na$_2$SO$_4$. C_{dl} and constant phase element (CPE) of the Gr/copper samples suggest an electrochemically smaller active surface to interfacial charging (discharging) and adsorption (like a non-Faradaic Helmholtz capacitor) as the density of conductive pathways decreases besides the lower effective surface roughness than that of the mechanically polished substrate. The CPE was 2.7 times higher than that
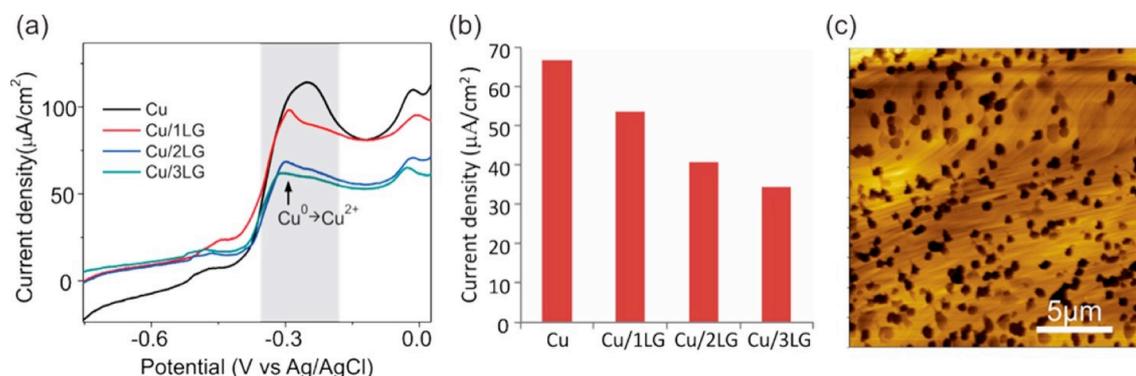
Figure 5: CVs of (a) bare Cu and Cu passivated by 1-3 layers of Gr, (b) intensity of oxidation currents under various passivation conditions, (c) atomic force microscope image of Cu/Gr after etching (adapted with permission from Ref. [47] Copyright (2014) American Chemical Society)
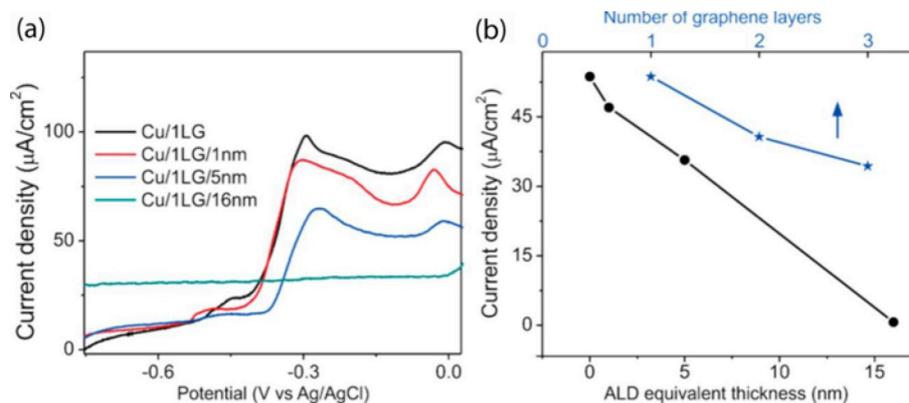
Figure 6: CVs of (a) Cu/Gr/Al samples, (b) Peak currents of Cu(0) for Cu/Gr/Al with different aluminium thicknesses and in comparison to Cu/Gr with 1-3 Gr layers (adapted with permission from Ref. [47] Copyright (2014) American Chemical Society)

measured in NaCl solution (0.1 M) and over ten times lower than in Na_2SO_4 solution (0.1 M). The resistance of coated copper to corrosion (R_{corr}) was ca. thirty five times higher than those of mechanically polished ones 6×10^4 and 1.6×10^3 Ω cm², respectively. R_{corr} was appreciably lower than that of NaCl and higher than with Na_2SO_4 solutions. Thus, inhibition efficacy was found as follows; >97% in 0.6 M and >71% in 0.1 M NaCl. Thus, performance was sensitive to the integrity of the Gr film as mechanically polished samples indicated rapid protection efficiency losses with increasing chloride concentrations in the test solutions.

Rigorous corrosion testing of nickel/CVD grown Gr films revealed low intensity oxidation of the substrate by heat treatment. Oxidation resistance was effective up to 500 ºC for 3 h, and with durable exposure (2 h) to essential hydrogen peroxide (31 wt.%) solution [49].

Both nickel and copper were coated with ML Gr films [14] by mechanical transfer. CV measurements indicated a remarkable restriction of the evolution of anodic and cathodic current transients, resulting in the effectively decreased electrochemical activity of the metal substrates. *Fig.7* depicts CVs of bare and Gr-coated copper samples, and XPS spectra of copper under ambient conditions, at 100 mV and sputter cleaning. Region (i) and (ii) contains a shoulder and a peak due to CuO and Cu_2O/Cu, respectively. The third part of the figure shows SEM images of Cu and Gr/Cu before and after voltammetry scans. Based on Tafel

analysis, corrosion rate assessment (quantification prediction) for bare substrates gave the following results: ca. 6×10^{-13} and ca. 3×10^{-14} m s⁻¹ for copper and nickel respectively. Graphene/copper indicated lower corrosion rate of ca. 8×10^{-14} m s⁻¹ up to ca. seven times less than the bare substrate. The Gr/nickel showed a reduced rate of ca. twenty times to 2×10^{-15} m s⁻¹. With two or four folds of Gr layers, the corrosion rate of nickel decreased to 2×10^{-14} and 8×10^{-15} m s⁻¹ respectively, which is considered as a notable performance improvement.

EIS measurements suggested an undamaged state of Gr, but the metal surface might corrode at high rates around void areas in the film. The most probable electrical network included resistive elements of the solution; $R_s \approx 40$ (6.5 Ω cm²) and charge-transfer at the interface; initial $R_{ct} \approx 3 \pm 0.5$ kΩ cm² due to Faradaic processes. The mass transport process was modelled with a Warburg element ($W \approx 1.0 \pm 10^{-3}$ Ω^{-1} s$^{0.5}$ cm⁻²) as a semi-infinite diffusion component. The inhomogeneity of the surface was handled by incorporating a CPE as a non-ideal capacitance in the circuit due to the impedance dispersity of the electrical double layer component at the metal-liquid interface. Double layer capacitance was best achieved with 2.2 μF cm⁻² and the CPE was $Y_o \approx 7.6 \times 10^{-5}$ Ω^{-1} s cm⁻². Increased resistance coupled in series was caused by the electrical surface resistance of Gr ($\approx$1 kΩ) with a double layer capacitance of ~3.8 cm⁻². Average fitting parameters were changed
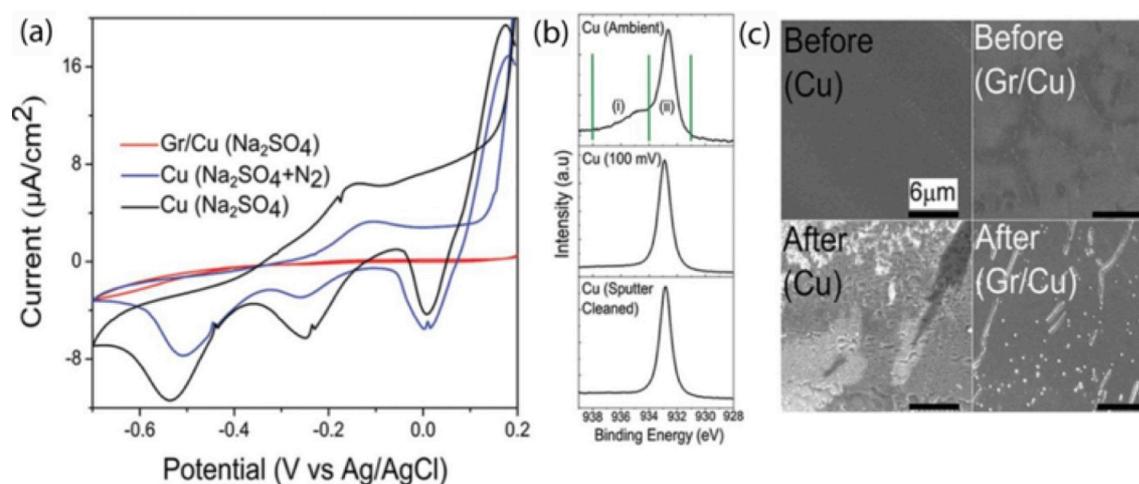
Figure 7: CVs of (a) bare Cu and Gr/Cu samples with the blue line corresponds to the measurement with N_2 bubbled through the solution; (b) XPS spectra of copper under ambient conditions at 100 mV and sputter cleaning with regions (i) and (ii) are due to CuO and Cu_2O/Cu(0), respectively; (c) SEM images of Cu and Gr/Cu before and after CV scan (adapted with permission from Ref. [14] Copyright (2014) American Chemical Society)

accordingly as follows: $C_{dl} \approx 9\pm5.0$ μF cm^{-2}, CPE: $Y_o \approx 2\times10^{-5}$ Ω^{-1} s cm^{-2}, W $\approx 3\pm0.5$ Ω^{-1} $s^{0.5}$ cm^{-2}, and $R_{ct} \approx 10\pm1.2$ $k\Omega$ cm^{-2}. The C_{dl} was ca. ten times lower than measured with bare copper, which seemed to be unrealistic. Such a difference is thought to originate from the quantum capacitance of the Gr film as a contribution in series with the C_{dl}. This is partly due to the small density of states in the Gr, which is in good agreement with the reported quantum capacitance of 7–10 μF cm^{-2} [50]. The increased R_{ct} of the copper samples in comparison with the one measured at the beginning of corrosion test suggested that the minor proportion of the metal substrate was uncoated with Gr, participated in corrosion processes. The proportion of this unburied area was estimated to be unexpectedly high ($A \approx 0.3$), exceeding the scale impressed by microscopy images ($A < 0.05$). Such a difference was attributed to the access to larger substrate areas by the diffusing depolarisator species and a subsequent greater proportion of Helmholtz capacitance then pseudo-capacitance based on electrochemical activation of metallic substrates for anodic reactions.

Nevertheless, the ability of Gr to catalyse, and depolarise cathodic reactions such as oxygen reduction [51] by well-coupling micro-galvanic cells in contact with the poly-aromatic sheets [52] either in planar or concentric forms (according to the function of the surface quinone group on carbon electrodes, i.e. GARTEN and WEISS's mechanism) is almost always a concern. Thus, protection efficiency depends largely on the deposition technique, quality, and epitaxial pile structure of the Gr sheets, affecting ionic and electrical transport through the films. Due to low adherence as a result of the lack of direct grafting or anchoring type chemisorption, direct CVD-grown films are classified as fragile coatings liable for potential polarisation and charging of the surface due to their structural alterations [53]. To increase the small domain size of the flakes, high temperature annealing at ~1035 °C on the metal surface offers a resolution promoting further development of coherency of the films composed of large domains [54]. However, there are many

problems with direct CVD deposition on metals like nickel because of the film formation mechanism of polyaromatic carbon materials.

The mechanisms of surface catalysis and adsorption during Gr growth on copper were determined [13, 55] to be different from film growth on nickel, which is governed by the combination of catalysis and adsorption as well as solubility and rejection. As the CVD technique enables the synthesis of Gr on larger metal surfaces [22, 56], carbon sources are thermally decomposed and fragmented into smaller segments, then atoms arrange into 2D honeycomb lattices. The preparation of CVD grown Gr films takes place readily on surfaces with face-centred cubic and hexagonal close-packed crystal structures. Methane and ethylene, which decompose at a temperature of ~1000 ºC are used as main precursors [24, 25] even though such a high temperature might significantly affect the metallographic state of the substrate, nucleation and deposition kinetics. Metals with close to zero solubility of carbon at the reaction temperature, e.g., copper, deposit *via* a catalytic mechanism and they are more likely to assist the formation of SL Gr instead of ML ones owing to the lack of catalytic effect after the first layer is deposited. Thus, the metal surface acts as a catalyst facilitating Gr growth in the form of a finishing coating, as deposition is restricted to SL [57]. Film growth on metals with mediocre carbon solubility at reaction temperatures undergoes *via* a complicated mechanism in which carbon diffuses into the bulk metal, e.g., nickel, and is rejected during the cooling process because of the decreasing solubility at lower temperatures. The rate of carbon rejection is critical to the development of ML Gr coatings. Hence, these films form preferentially instead of SLs on copper. Further complications may arise by the use of thicker metal specimens because thicker substrates dissolve larger amounts of carbon during the cooling phase since they release more carbon to produce larger amounts of carbon soot. Although any process that needs to be scaled-up must be optimised for engineering alloys, there are many results of theoretical interests as well.

A thesis [58] thoroughly investigated the protection efficacy of direct CVD grown Gr on copper, depending on flake size and uniformity, duration and temperature of the heat treatment. Key factors leading to firm protection performance both in terms of instantaneous and long-term functions are identified as to eliminate defects in the Gr sheets, flakes must build up a flawlessly continuous, coherent, and uniform film. Low nucleation and high growth rates must be maintained under low pressure to gain a low density of carbon spots (as probability of deposition decreases) and coherent films with the absence of unburied areas at flake boundaries. The films with large flake sizes were tested in a number of experiments to optimise growth parameters in order to reduce the amount of disorders, defects and void areas. A series of chemical oxidation experiments were performed by heating bare and coated copper foils at various temperatures and for various time frames. Ellipsometry helped characterise the thickness of oxide layers on the highly reflective copper. Heat treatment was carried out at 120 ºC under air for up to 76.5 hours. The growth rate of copper oxide layers, as an inherent indicator of the oxidation rate, almost linearly increased with time and the difference was only ca. a factor of six between the bare and coated copper foils. Corrosion deterioration was located at two spots where air leaked through defects in the Gr film and around its flake boundaries. Following growth time in low pressure CVD experiments, films were grown in a way that their flakes were not connected. Therefore, flake growth stopped before expanding sufficiently to connect with one another. Then oxidation was performed at 190 ºC for up to 37 h and 26 min. Optical microscopy images revealed a remarkable difference between the bare and coated copper samples. Then oxidation was repeated with closely spaced Gr flakes investigated by optical and SEM. The samples with closely spaced flakes gave very good results after a 30 min treatment at 190 ºC. A 13.5 h treatment led to minor deterioration in the substrate exhibiting black oxide patches on the surface but a 23 h and 35 min annealing resulted in corrosion of nearly the entire surface.

In a series of experiments, copper substrates coated with continuous Gr flakes were treated at 190 ºC for various periods of time. After a 30 min heating, samples exhibited a fine condition with some proportion of the surface composed of dark copper oxide. After a 10 h and 45 min treatment, migration of the corrosion front on the copper surface (CuO formation) proved to be far advancing and the heavily corroded areas showed a homogeneous distribution in the form of tree roots. Optical microscope images after 35 hours of heating indicated heavy corrosion, but the surface was not fully covered with oxides. Raman spectroscopy quantitatively characterised copper oxidation, which could proceed at flake boundaries as they feature the least hindrance to mass transport of depolarisators and the large sheets did not damage or etch oxidatively inside.

Apart from growth related issues, electrochemical aspects of the shortfalls may also be relevant. Gr films, acting as physical separators, might take part in electron transport, helping interfacial charge-transfer perpendicular to the surface based on good coupling and ballistic electron transport parallel to the basal plane of the sheets. Oxygen adsorption and reduction reactions with low activation energies may further complicate the application of this material, which has been recently underlined in a short- and long-term performance test of Gr/Cu and Si substrates [59]. Results clarified that the samples heated at 250 ºC for 6 min showed heavy oxidised spots on bare copper, but little oxidation on the Gr coated one. Therefore, Gr served as a sufficient oxidation barrier helping to preserve the metallic state from chemical oxidation over a short time period at high temperatures. Nonetheless, long-term annealing of coated copper led to severe oxidation probably because oxygen diffusion proceeded through discontinuities and defects around Gr sheets. After 17 h at 250 ºC, coated copper became heavily oxidised in a way indistinguishable from the bare specimens subjected to the same conditions. Similarly, a 15 min treatment at 185 ºC resulted in partial oxidisation of bare copper but coated samples remained almost entirely in a pristine state. Nevertheless, after 17 h annealing in air at 185 ºC, both bare and monolayer coated substrates became severely corroded and the products were identified as cupric oxide (CuO) in a stoichiometric ratio. The timescale of protection against thermal oxidation by Gr on copper was estimated to be ~1 h at 185 ºC and ca. ten times faster at 250 ºC. To explore the long-term protection ability under less severe conditions, bare and coated copper specimens were stored under ambient conditions (at 25 ºC) at low but variable humidity levels for up to two years. Bare copper showed relatively slow oxidation at room temperature. Gr/Cu tested after up to eighteen months exhibited oxide-free condition over the first two weeks but within a month the surface started to tarnish non-uniformly (in several hundreds of micrometres in size). Within a few months, most of the regions became heavily oxidised with an estimated oxide layer thickness of tens or hundreds of nanometres. Alignment between the substrate grains and Gr flakes was concluded to have an influence on oxygen and water diffusion along the surface. After eighteen months, the entire surface had become oxidised as EDAX analysis revealed a predominant percentage of cuprous oxide (Cu₂O) in the oxide phase. Gr/silicon gave similar results as bare and coated samples were compared to each other after one week under atmospheric conditions, indicating an increased oxygen content of silicon as time of exposure increased. According to the outcome of the series of experiments, short-term oxidation testing reflected some degree of protection but long-term propagation unveiled extensive wet corrosion, similar to that seen in the case of bare copper specimens. The high rate of the wet corrosion of copper is due to promoted galvanic cell formation on a micrometre scale at ambient temperature, especially over long-term exposure.

As a special field, the majority of medical applications require metals even today, in bone and joint replacements [60, 61], stents [62, 63] dental materials [64, 65], pacemakers and generators [66, 67]. Stainless
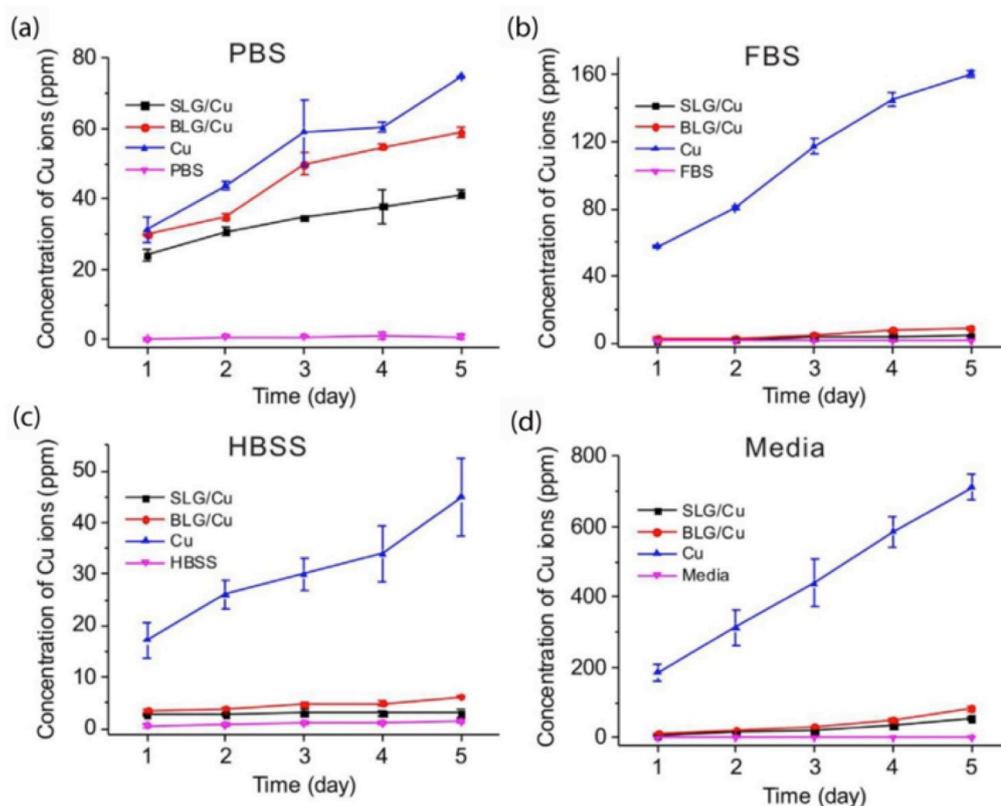
Figure 8: Concentration of Cu(II) ions from different biological solutions: in (a) PBS, (b) FBS, (c) HBSS solutions, and (d) cell culture media in the presence of SL (black) and BL Gr/Cu (red), Cu (blue) and blank solutions as controls (pink) measured by inductively coupled plasma-mass spectrometry (ICP-MS). All results are presented in a mass ratio (ppm) [79] (republished by courtesy of the publisher, Nature Publishing Group)

steel, titanium and cobalt alloys are the most common orthopaedic materials for joint prostheses articulated with a plastic bearing surface for many body parts [68, 69]. There are areas that require the use of alloys of mercury, silver, tin, copper along with limited amounts of zinc, palladium, indium, selenium and titanium in dental surgery [70, 71]. Despite the numerous successes, the main disadvantage of most of the alloys in medicine is their corrosion and the resulting physical deterioration (disintegration), quantity loss and contingent toxic side effects. The increased concentration of the aforementioned elements as soluble metals surely leads to chemical reactions of altered enzyme functions. It has been proven that soluble metals released because of corrosion processes can induce an innate monocytes-macrophage response and trigger immune responses, causing toxic, inflammatory, allergic or mutagenic reactions to patients [72]. In addition, corrosion leads to the adverse formation of metallic debris; solid particles and inorganic or organometallic compounds, in the periprosthetic soft tissues causing metallosis [73]. Furthermore, the corrosion degraded structural integrity leads to the premature failure and loosening of the metal devices [74]. Therefore, corrosion protection of metals used in physiological environments is of paramount importance and needs to be properly addressed for the sake of durable successful biomedical applications. Currently ceramic coatings are the most widely used to isolate metals from body fluids, e.g., zirconia transited from a zirconium alloy minimises corrosion in knee

implants but the implants roughen with time. Nano-size diamond coatings and apatite-nano-diamond composites are promising candidates for the protection of metallic implants [75, 76], but they are highly permeable and greater surface roughness notably decreases their performance. Nowadays, atomic thin Gr films enhance the bio- and hemo-compatibility of implants to some extent [77, 78]. Interestingly, the outcome of almost all cell viability tests showed the biocompatibility of Gr on the nickel-titanium surface with some protection. After considering these facts, studies focused on clarifying the possibility of application of Gr as a corrosion protective film on structural metals used for biomedical devices. The most important future application would be the efficient corrosion protection of metallic prosthetics to avoid serious health problems to patients. In an effort to offer an alternative to protect metallic implants, Gr was proposed as a biocompatible and protective film as well [79]. Copper was the substrate for cell viability tests to reveal any change in immune response (lymphocyte transformation test) and the metal sensitivity test because of its high toxicity [80] making the experiment sensitive to reflect any significant protection. In vivo experiments served to study the protection of copper in live animals to evaluate Gr as a biocompatible film in physiological conditions. Corrosion inhibition of copper with Gr films was studied by the way of release and accumulation of copper ions in biological aqueous environments of phosphate buffered saline, fatal bovine serum and Hank's balanced salt solution. Cell culture
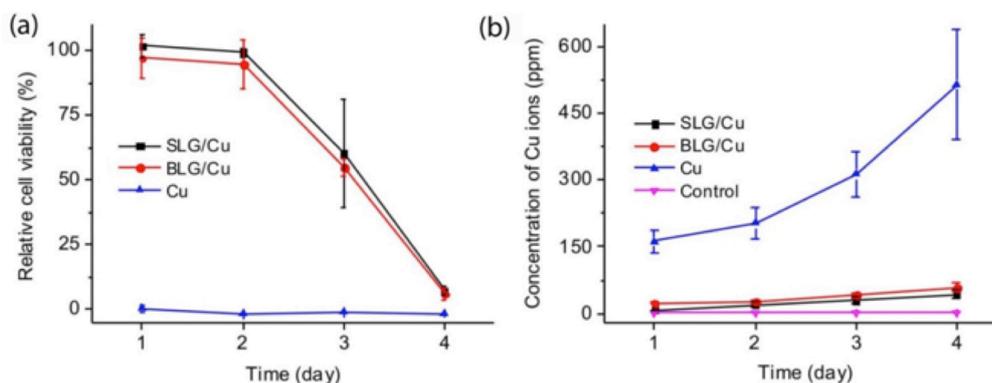
Figure 9: Relative cell viability and concentration of Cu(II) ions as a function of time. (a) Relative cell viability vs. time for SLG/Cu, BLG/Cu and Cu; (b) Concentration of Cu(II) ions in the cell culture medium after incubation with SLG/Cu. BLG/Cu and Cu (calculated from ICP-MS). A medium taken from a regular cell culture without Cu foils was used as the control [79] (republished by courtesy of the publisher, Nature Publishing Group)

media were monitored with inductively coupled plasma optical emission spectroscopy (ICP-OES) for five days. Single-layer Gr/copper indicated slightly varied but very good conditions on average.

The copper ion release rate was found to be at least an order of magnitude lower compared to bare copper that correlates almost linearly with time. The copper concentration in the solutions was higher when bilayer Gr/specimens propagated as compared to those of the tests with SL Gr obtained in all solutions. This was attributed to the stronger adhesion of SL films to metallic surfaces than bilayer ones, probably because of the better adaption to surface topography with high flexibility [81]. *Fig.8* shows the concentration of copper ions detected in different biological solutions such as phosphate buffered saline (PBS), fetal bovine serum (FBS), Hank's balanced salt solution (HBSS) and cell culture media with SL and BL Gr films and blank solutions as controls. Cell viability tests indicated almost the same results with single- and bilayer Gr/copper samples showing ±2.2 and 2.6% differences respectively compared to the control samples after the first day. Nonetheless, the results produced by low concentrations of dissolved copper in test solutions do not mean good results overall because the ion concentration could be much higher in a micro-milieu between the cells and coated substrate. The point of such a difference can explain why the cell viability dropped remarkably by more than 90% over the second, third, and fourth days of the test compared to the Gr contained samples although the copper concentration in the solutions remained very low (less than 70 ppm). Probably interfacial accumulation of metal ions must play a pivotal role in exceeding the toxicity limit, inevitably triggering lower cell viability. To resolve this problem, a further modification of the metal surface was necessary [82, 83]. By self-assembly monolayer treatment with decanethiol, bare and Gr/copper was modified with decanethiol, which was obviously assembled on the bare surface around the boundaries of Gr flakes. This is because the seemingly proper protection of copper was accompanied by much improved cell viability over three days *(Fig.9)*. The thiol modified SL Gr/Cu indicated that cell viability was ca.

55.8% after an incubation period of three days compared to SL Gr/Cu *(Fig.10)*. Thiol modified copper showed an improvement of up to 2% in comparison to the less than 1% viability observed for bare copper. These results suggest that Gr on copper is the major factor for protection, which can be appropriately enhanced by thiol modification. The film was also found to reduce immune response to copper in a clinical setting by a lymphocyte transformation test.

Animal experiments indicated a positive outcome to coated copper samples tested under *in vivo* conditions. The copper concentration was less than 3 and ~4.5 ppm in blood samples extracted from living rats implanted with SL coated and bare foils, reflecting the potential of thin Gr films in biomedical applications.

CVD-Grown Graphene Loading on Metal Surfaces by Mechanical and Electrical Techniques

Spin-Coating

There are a number of techniques to obtain high quality thin films. Static and dynamic dispense techniques such as dip [84-86], angle-dependent dip [87, 88], and hot dip spin, spin [89-92], spray [93], flow [94, 95], capillary [96, 97], roll [98], and reverse roll coating [99, 100] have developed thin layers on solid surfaces. By careful design, it is also possible to complement one another, offering a way to be tailored to application demands. Optimisation of these techniques is mostly dependent on actual parameters of the fluid phases and molecular or colloid solutions such as concentration, temperature, pressure, partial pressure or volatility of substances and solvents, dynamic or kinetic viscosity, and surface tension of fluid phases. The texture (roughness) of the substrate surface has much less influence (except for the surface energy) on the quality and quantity of casting. Drying and curing might later take place. From them, the ones featuring well controllable Gr growth are detailed in the following. Both spin-coating (SC) and kinetic spray (KS) are relatively novel techniques enabling us to form coatings of good quality on any sort of surface texture regardless of the substrates. SC is
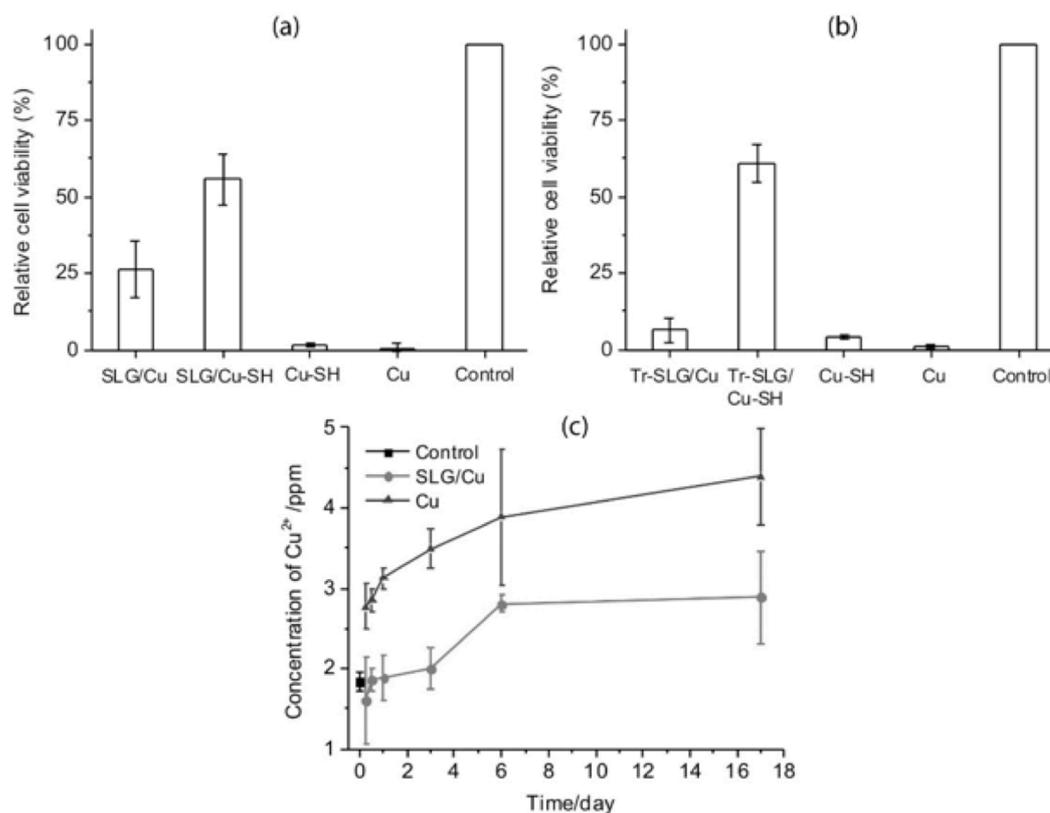
Figure 10: Relative cell viability of (a) bone cells incubated with SL Gr and SL Gr/Cu-SH, Cu-SH and bare Cu foil for 3 days. Control is a regular cell culture without Cu foils; (b) Relative cell viability of bone cells incubated with Tr-SL Gr/Cu, bare Cu, Tr-SL Gr/Cu-SH, and Cu-SH for 1 day; (c) Concentrations of Cu(II) ions in blood samples extracted from live rats with concentrations of Cu(II) from normal rats before implantation as control (black square), rats with implanted SLG/Cu foils (blue triangles) and rats with implanted Cu foils (red dots) as a function of time [79] (republished by courtesy of the publisher, Nature Publishing Group)

suitable to work out and implement practical deposition methods for atom-thick films on various metals, although a different approach must be applied for Newtonian [101] and non-Newtonian fluids [102]. Uniform coatings can be obtained over large areas with reproducible thicknesses in a good controllable manner, but only on planar substrates, which is a major limitation of this technique. SC starts by applying a small puddle onto the centre of the substrate, then spinning commences at initially a lower then, within a matter of seconds, at a higher rotational speed (two stages of centrifugation up to ~3000 rpm). The centrifugal force and centripetal acceleration result in excessive spreading of the fluids on the surface (wetting and covering depend on multiple physical-chemical parameters) and an inevitable loss (can be recycled) around the edges of the substrate until the spin-off time is completed. Then, a thin film forms releasing solvents and curing takes place, because of chemical reactions.

Kinetic Spray

Kinetic spray (KS) utilises a Laval nozzle to spray Gr colloid solutions at a very high speed and the supersonic acceleration leads to the production of very small droplets that evenly disperse then evaporate reducing the tendency of flake aggregation. Besides the uniform and smooth layers, defects in the Gr flakes were not observed. Hence, high quality film production by KS depends on the energy of the impact stretches Gr sheets then carbon atoms arrange into flawless hexagons based on plasticity probing of the Gr flakes [103]. Transmittance of graphene oxide (GrO) film spray deposited on soda-lime glass (from a 10 wt.% colloid solution of GrO in ethanol) was 73% and sheet resistance was 19 kΩ sq^{-1} using incident light at the wavelength of 550 nm. The electrical conductivity of the coatings increased by up to a thousand times after hydrazine reduction and annealing at 400 ºC. Additional thermal annealing at 1100 ºC further enhanced conductivity by ca. 10 times. Soda-lime glass with a transparency of 80% and electrical conductivity of 209 S cm^{-1} (or sheet resistance; R_s = 2 KX sq^{-1}) as applied to a 14 nm thick Gr film [104]. KS is a reliable technique of wide commercial availability and offers a flexible solution to large-scale using according to recently reported corrosion tests [105-107].

Electrospinning

Electrospinning (ES) creates polymer fibres with a diameter range of between 40 and 2000 nm. The fibres can be made from solutions or from fused materials, controlling diameter size through the adjustment of

surface tension, solution concentration, conductivity and so forth [108-110]. ES occurs when the electric force of the surface overcomes the surface tension of the fluid phase triggering an electric spark, results in the solution being expelled from a syringe forming jet flow impacts and deposits. When the expelled material dries out or solidifies, it forms an electrically charged fibre that can be directed or speeded up by electric forces. So, a polymer solution stored in a syringe is charged to a high electrical potential. As the jet stretches and dries, radial electrical forces cause it to splash repeatedly. Dried, solidified fibres are collected on an electrically conductive (metal) screen. By ES, the composition of nylon 6-6 and Gr was prepared and deposited on porous silica then tested under various conditions [111]. Oxidation and reduction potentials in acidic, alkaline, and peroxide solutions were assessed by dynamic polarisation. SEM observations determined that treatment with acidic and basic solutions led to the formation of solid aggregates and slightly folded GrO flakes. Thermal treatment at 700 ºC gave samples with a porous structure. After annealing, a subsequent chemical derivatisation with acidic, basic and peroxide solutions assisted with ultrasonication facilitated the separation of GrO sheets, leading to a solid composed of lower dimensional flakes, especially in the case of oxidation. Oxidation was carried out in KOH and the reduction in H_2SO_4 solution. EIS data obtained in the immersion test with a Na_2SO_4 solution indicated decreasing interfacial capacitance of the ES produced samples with the Ny-GrO coatings. The EIS data calculated at 0.8 Hz changed between $1.4–4.5×10^{-5}$ and $7.5–8.9×10^{-7}$ F cm^{-2} (depending on the concentration of the electrolyte) with the GrO content increasing from 0.4 wt.% to 2 wt.%.

Electrophoresis

Electrophoretic deposition (EPD) is a promising technique to fabricate functionally graded [112-114] and hybrid composites [115, 116], laminated nano-ceramics [117, 118], and functional, nanostructured films and coatings [119-121]. In the last couple of years, numerous EPD applications have emerged as industries count on commercial advantages over other fabrication routes. It is a versatile and cost effective technique with acceptable levels of control over microstructure, stoichiometry, and macroscopic-microscopic dimensions and properties [122, 123]. EPD is a two-step colloid process in which electrostatically charged particles, suspended in a liquid medium, migrate according to the electric field towards the oppositely charged electrode. Then particles deposit, and flocculate on the electrode surface forming a relatively dense, homogeneously packed, bonded layer. Increasing the volume fraction of the nanoparticles and their deposition on the surface of any electrically conductive material is relatively easy. Usually, a post-treatment is required to increase the density of deposits and eliminate porosity [122]. Sometimes special modifications are required to render surfaces with

notable properties such as increasing hydrophobicity by silylation. EPD is performed in stable colloid solutions in which there are plenty of electrostatically charged particles migrating in response to an impressed DC electric field and deposit either as a loose homogeneous or compact film onto the oppositely charged electrodes. To make deposits denser and eliminate porosity, additional EPD or heat treatment is required. Oxide particles, conductive polymers [124] and highly dispersed activated carbon [125] of nearly any size in colloidal suspensions [122] were deposited by this method, reflecting its high versatility. Furthermore, by electrophoresis uniform thin layers can be made on highly ragged, complex surfaces. Moderate electrical conductivity of the resulting substrates is the only requirement. The sample surface is treated as a whole in a uniform manner, whereas dimensions, deposition rate, uniformity and scale up features are also favourable. EPD is a preferable alternative to many other methods like slurry dipping, thermal and plasma spraying, sputtering, and physical/chemical vapour deposition.

In the course of thick Gr layer deposition on copper by EPD [126], deoxygenation of carbon was hinted at by bathochromic shifts in the UV spectra because of the electrochemical configuration and reaction mechanism. Potentiodynamic scanning showed positively shifted corrosion potentials and lower corrosion currents (I_{corr}), which suggested only mediocre protection by the thick Gr films in the initial phase of immersion propagation.

The composite coating of reduced GrO was prepared by cathodic EPD in an aqueous solution in a symmetrical cell-electrode arrangement [127]. Optimum conditions for cathodic deposition were found to produce a thickness of ~40 nm by applying 10 V for 30 s. SEM observations revealed the size of the GrO sheets being 1–2 μm, covering the surface uniformly. The composite coating was shown to protect copper firmly from electrochemical corrosion. Potentiodynamic measurements indicated an ennobled free corrosion potential with slightly reduced anodic and cathodic current branches, suggesting a lower corrosion rate of the composite coated copper compared to its bare form during the early phase of propagation. Experimental conditions are summarised in *Table 2*.

By electrophoresis, GrO was uniformly deposited on a permanent magnet, e.g., NdFeB and the coating was subsequently reduced to partially remove functional groups containing oxygen [128]. The EPD GrO coating showed substantially better adhesion to the substrate compared to the spin-coated ones. The surface of the bare NdFeB was rough and porous but site-dependent topography of the EPD-GrO modified samples depended on the amount loaded. The relationship between the increasing thickness and decreasing adhesion of the coating with an extension of deposition time was determined. Raman, infrared, and X-ray photoelectron spectroscopy (XPS) results reflected the strong adhesion of EPD films and that was connected to the Koble-type decarboxylation mechanism [129, 130] of the surface yielding GrO flakes. An unexpected outcome was that several deterioration reactions were observed in GrO. The signal of epoxy groups decreased

Table 2: Metallic substrate, type and growth parameters to obtain Gr film by EPD and chronoamperometry

Substrate	Preparation types and their parameters	Corrosion test parameters	Work-electrode surface (cm^2)	Types of potentiodynamic tests	Ref.
Cu	GrO, EPD in sol. C = 1.0 mg cm^{-3}, E=1 V / 10 mm^{-1}, 10 min	air ox. at 200 ºC for 4 h, in 30 wt.% H_2O_2 for 2 min, then 0.1 M NaCl solution	Unknown	after 1 h at OCP, scan rate: 1 mV s^{-1}	[131]
Cu	red. GrO with polyiso-cyanate cured with HAA (hydroxyl acrylic acid), EPD for 30 s, at E = 10 V 10 mm^{-1}, 0.1 M $NaBH_4$ red. for 5 min	3.5 wt.% NaCl, at 25 ºC	1	after 2 h at OCP, scan rate: 1 mV s^{-1}	[132]
NdFeB	GrO red., EPD in sol. of 1 mg cm^{-3} GrO, NdFeB anode & Pt cathodic dep., E=10 V / 10 mm^{-1},	3.5 wt.% NaCl solution at room temperature	Unknown	EIS	[133]
Cu	GrO isocyanate cured with hydroxy acrylic, then silylation, EPD on Cu at 10-30 V 10 mm^{-1}	3.5 wt.% NaCl, at 25 ºC	1	EIS	[136]
Cu	chronoamperometry of GrO at −1.5 V for 10 min in sol. C = 2 g dm^{-3}, in 0.1 M KCl, at −1.5 V for 10 min, then GrO red. $NaBH_4$ (0.1M) for 5 min	OCP 3.5 wt.% NaCl sol. after 1 h	1	EIS	[138]
Mild steel	EPD in Ni sol. at pH = 3, Iρ = 1 A dm^{-2}, at t = 40ºC, Gr c=100 mg dm^{-3}	3.5 wt.% NaCl, at 27 ºC	1	scan rate: 0.01 V s^{-1}, EIS	[169]

after deposition, which might be explained by the high anodic electrode potential applied to the working electrode and the large potential drop at the interface. The decrease in current density during potentiodynamic scanning and the positive shift in the free corrosion potential of the system reflected efficient initial protection but lower cohesion and adhesion of the thicker coatings, leading to a poorer protection potential.

The GrO-acrylic polymer composite fully covered the copper substrates, but the sheets were separated far from one another in an island like manner consequently a large percentage of the surface remained unburied [131]. As for the reagents, solution and cell arrangement, EPD conditions e.g. voltage (10 V) and deposition time (30 s) proved to be optimal to obtain crack-free films of an average thickness of 45 nm. An electrochemical investigation indicated insignificant protection performance of the composite coating under immersion propagation. A steady corrosion current extrapolated to the OCP region showed slight changes over the anodic and cathodic regions. A current density drop of 38.3–3.5 μA cm^{-2} was observed in comparison to the bare copper. A polarisation resistance of the composite covered specimen was assessed by EIS, assigning an increased charge-transfer resistance of ~25 Ω cm^{-2} (bare copper: ~8 Ω cm^{-2}).

Stainless steel specimens with an exposed geometrical surface of 100 cm^2 were almost uniformly deposited with GrO in H_2SO_4 solution by coupling the substrate cathodically [132]. For transferring purposes, chemical and electrochemical etching was developed to delaminate the reduced GrO film, giving coherent freestanding membranes or deposit them onto other substrates. By optimising electrophoretic parameters, a low voltage of 3 V was found to consolidate reduced GrO layers preferentially aligned in an in-plane direction through a cohesive electrophoretic squeezing

force in near volume range of the electrode (cathode). The free-standing Gr membrane was reduced, annealed at 1000 ºC then a graphite-like architecture evolved with a *d*-spacing of 3.42 Å and C/O ratio of 16.7. Electrical conductivity was of high as 5.5×10^5 S m^{-1}.

Graphene Coated Metals in Electrochemical Cells and Batteries

Because corrosion is a serious problem in energy storage devices like batteries and fuel-cells [133], the application of GrO as a barrier coating to inhibit the corrosion of aluminium electrodes in lithium ion batteries is considered to be an innovative solution. Spin-coating was employed to coat aluminium with GrO and decrease the corrosion rate when the electrode was subjected to the electrolyte of $LiPF_6$ (1.0 M) dissolved in ethylene carbonate and dimethylcarbonate (50:50 vol.%) [134]. SEM observations with energy dispersive X-ray spectroscopy analysis and atomic force microscopy (AFM) scanning suggested increasing surface roughness of aluminium with lower GrO quantity, but its greater amount helped reduce the roughness remarkably. The average roughness of the bare substrate was around 350 nm, which changed to 199-154 nm for the GrO loaded one. The thickness of the film showed great epitaxial variety even though the coating was relatively thick. Cyclic voltammetry showed a decreasing electrochemical current response of the GrO coated electrodes with increasing surface loading. This indicated hindered and impeded charge-transfer through the electrode interface without any direct implication of capacitive and adsorption charge loss in the current transients. Accordingly, chrono-amperometry curves exhibited moderately changing characteristics (as a consequence of altered electrode kinetics) in the early phase of the transients and

decreasing base currents over the complete timescale because of the smaller electroactive surface. EIS spectra suggested a somewhat higher resistance against pitting, the lower susceptibility to localised corrosion. Thus, GrO restricted aluminium oxidation, behaving as a charge-transfer barrier. Otherwise, the lower capacity and potential of the cells should have been measured by the set-ups with the GrO loaded electrodes during charge–discharge testing. Instead, surprisingly the opposite was reported. Nevertheless, enhanced cycle-life stability (lower rate of capacity loss) of the modified electrode compared to the cell configuration with bare aluminium is of course self-explanatory. Self-discharge was lower as a decreasing rate of open circuit voltage decline was observed to the modified aluminium samples. GrO films did not show any corrosion and the underlying native oxide layer remained intact under test conditions.

Microbiologically induced corrosion (MIC) severely limits the structural integrity and lifetime of metallic structures in technological processes and applications because of frequently evolving mechanisms, i.e., crevice and pitting local corrosion phenomena. A microbial fuel cell (MFC) is a galvanic cell producing an electric current that causes bioelectro-chemical oxidation of organic substances on anodes and abiotic reduction on cathodes [135]. MFCs have been used as a galvanic tool to simulate extracellular electron transfer mechanisms of microbes and the bio-electrochemical oxidation of organic substances. In addition, metals are not used as anodes in MFCs due to their fast galvanic corrosion. Owing to this, Gr films are employed as barrier coatings and the corrosion rate of the working nickel electrode was evaluated by galvanic coupling in the MFC set-up [136]. Regarding the experiment, it was determined that the geometrical ratio of the working and counter electrodes was the guarantee to maximise, and concentrate current densities on the working electrode.

Nevertheless, the cell design did not allow the assessment of adequate kinetic parameters (general corrosion rate) because of the large uncompensated ohmic drop and unfavourable cell geometry. These are all relevant in DC techniques, although inadequate voltammetric data could provide a rough estimation, even though 5 mV s^{-1} was higher than recommended and might have resulted in the considerable rate of adsorption and pseudo-capacitive current during the measurement. In relation to the impedance measurement and its evaluation, these data are even more sensitive to inappropriate cell design. The protection efficacy of the Gr film was assessed under realistic operating conditions, the bare and coated nickel electrodes represented a 3D microporous structure. The overall outcome confirmed a retardation of corrosion processes by the Gr film for extended periods. CV scanning showed decreased current transients of coated nickel, which was ca. ten thousand times lower than measured for the bare nickel electrode based MFC after a propagation for 2800 h. Nickel dissolution rates of the Gr coated anodes were at least ten times lower than the baseline (uncoated) electrode. EIS characterisation confirmed impeded MIC on the Gr/nickel for ca. 40

folds in comparison with uncoated electrodes. The effective masking of nickel from the mixture of microorganisms and their metabolic products was concluded.

Composite Graphene Coatings

Physically Mixed and in situ Polymerised Mixtures

Chronoamperometry was successfully employed to prepare Gr coated copper, starting from an aqueous colloid solution of GrO [137]. As a result of the EPD process, a seemingly compact and void of area without coating; and polymer film with Gr inclusion was obtained. Nevertheless, the distribution of Gr flakes was highly uneven and inhomogeneous (random), whereas the sheets showed various degrees of folding and overlapping in the polymer matrix. In addition, the fast reduction of GrO in the polymer proves the high permeability of the matrix, although it reached a thickness of several micrometres. On the other hand, the coatings with reduced GrO provided moderate protection for copper with an initial inhibition rate of 94% drawn from the potentiodynamic polarisation measurement data (at nil impressed potential). The corrosion rate reduced to 0.2 mm yr^{-1} from the rate for pure copper of ca. 3.65 mm yr^{-1}. This assessment was only made at the beginning of the immersion test (after an hour), whilst waiting for stabilisation of the freshly installed cell set-up. In comparison with the EIS investigation, the composite coating provided electrical insulation and ionic diffusion hindrance to the copper surface. So, the protection was related to the increased polarisation resistance of Gr loaded coatings of ca. 26 Ω cm^2 compared to bare copper of 9 Ω cm^2.

Intrinsically conducting polymers (ICPs) like electroactive polyaniline (PAni) have been widely used materials for some time providing corrosion protection regardless of the long-standing disagreement over the anodic protection mechanism and the generally less effective barrier nature. PAni is a preferable electron-conducting polymer with favourable properties over other alternatives. Under optimal conditions, the protection mechanism leads to the formation of a passive (dense) metal oxide layer [138-140].

To increase protection performance, many substances such as well dispersed clay [141] with an aspect ratio of ca. 200 was aimed at lengthening mass transport pathways through the composite films for all redox active species. These polymer-clay mixtures exhibited notable protection improvements over neat polymers. Besides the several attempts to combine of carbon nanotubes (CNTs) with organic [142-156], inorganic, and metal [157-160] matrixes for corrosion protection and resistance, respectively, there are examples from the literature about testing Gr-polymer composites for various purposes [160-162]. The main focus of these works was the exploitation of high relative aspect ratio of Gr (ca. 500) [163] in comparison to exfoliated clay minerals. Hence, Gr is a key filler in advanced gas barrier materials, motivating studies to

Table 3: Metallic substrate, type and growth parameters to obtain Gr films by spin coating and solution casting. Types and settings of corrosion tests and electrodes with the reference number of the article

Substrate	Preparation type and parameters	Corrosion test parameters	Work-electrode surface (cm^2)	Types of electrochemical tests	Ref.
porous silica	electrospin Gr-Nylon 6-6 in formic acid agitated for 12 h then add. of GrO, ES at 12 kV, flow rate: 0.2 mL h^{-1} for hours, Electrode: polymer solution 90 wt.% of formic acid with 0.36-2 wt.% GrO and Ny	Potentiostatic oxidation in KOH (0.5 M) for 3.5 h, reduction in H$_2$SO$_4$ at pH = 2 for 8 h, EIS in Na$_2$SO$_4$ in (1-10^{-1}-10^{-2}-10^{-3} M) for hours	unknown	potentiostatic treatment, potentiodynamic polarisation, scan rate: 100 mV min^{-1}, EIS	[112]
AlO$_x$	GrO sonication in ethanol then spin-coating	LiPF$_6$ in ethylene carbonate and dimethyl-carbonate (50:50 vol.%); at 1.0 M	100	cyclic voltammetry, chronoamperometry, potentiodynamic polarisation and EIS	[114]
Cu	Spin-coating of Gr with acetone (20 cm^3 cm^{-2}) then rapid annealing 800-1000 ºC for 3 min	Seawater 3.0–3.5% NaCl	0.25	potentiodynamic polarisation, scan rate: 2 mV s^{-1}, EIS	[136]
steel	PAni-Gr comp. (NMP complex and casting)	3.5 wt.% NaCl solution	unknown	potentiodynamic polarisation	[153]
cold-rolled steel	GrO red. with N$_2$H$_2$, pressing powders then curing (thickness: ~120 µm)	3.5 wt.% NaCl, at 25ºC	1	potentiodynamic scanning at 10 mV min^{-1}, after 30 min EIS	[161]
low alloy steel	2 wt.% Gr/MWCNT/PEI & 20 wt.% Gr/PEI comp.-s, dispersion in PAA cured at 150 ºC for 5 min, then 250 ºC for 5 min for imidisation	3.5 wt.% NaCl solution	9.62	potentiodynamic polarisation (scan rate: 1.67 mV s^{-1})	[7]
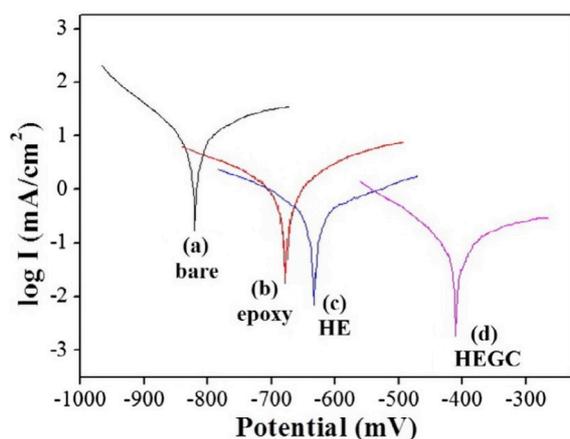
Figure 11: Tafel plots of the (a) bare, (b) epoxy-coated, (c) HE coated, and (d) HEGC-coated CRS electrodes [178] (by courtesy of the publisher, Elsevier)

achieve very low oxygen permeability using low filler contents (compared to neat polystyrene and the best polymer-clay mixture known at that time) [164] and underscore the possible obliteration of traditional clay-polymer composites [165-169]. Thus, a carefully designed and well-prepared PAni-Gr mixture exhibited outstanding impermeability characteristics against the diffusion of molecular oxygen and water vapour *versus* neat PAni and PAni-montmorillonite using a filler content of 0.5 wt.% and the excellent protection performance of steel substrates. Potential ennoblement of steel electrodes with coatings of increasing Gr content was considerable and an approximately fifteen times lower steady corrosion current was reported. The main reason for the success is mostly related to the high dispersity of Gr in the composite of microscopically isolated and macroscopically non-percolating

distribution [170]. Experimental conditions are summarised in *Table 3*.

Lower electrical and ionic conductivities always provide better corrosion protection. Therefore, metals coated with Gr filled composites of electrically less or non-conducting (practically insulating) matrices like epoxy (in the case of high dispersity) can behave as a super-hydrophobic interface [171-176]. These composites are characterised by a water contact angle of at least 150º making them repellent, resistant to water absorption [177] and their exceptionally good anti-wetting properties results in good corrosion prevention potentials.

The nano-casting prepared epoxy-Gr composite featured high hydrophobicity and provided excellent protection for cold-rolled steel (CRS) panels [178]. *Fig.11* shows Tafel plots of bare, epoxy, hydrophobic epoxy (HE) and hydrophobic epoxy-Gr composite (HEGC) coated CRS panels. The open circuit potential of the immersion tested steel panels coated with epoxy-Gr showed appreciable ennoblement of ca. 0.3 V and ca. ten times lower corrosion currents compared to neat epoxy coated steel. The much higher polarisation resistance was partly attributed to the more hindered mass transport through the Gr loaded coating (with less than half the oxygen permeability) and access to its interface on the fluid side as a result of the finely patterned structure of the outer surface. The reproducibility of the data is questionable because parameters were not representative to properly describe the kinetic processes of the system, proceeding at the metal-solution interface. As in other cases, electrochemical techniques gave instantaneous rate assessments (in the first 30 min of the tests). Thus, the reported data are not suitable for forecasting long-term corrosion rates.
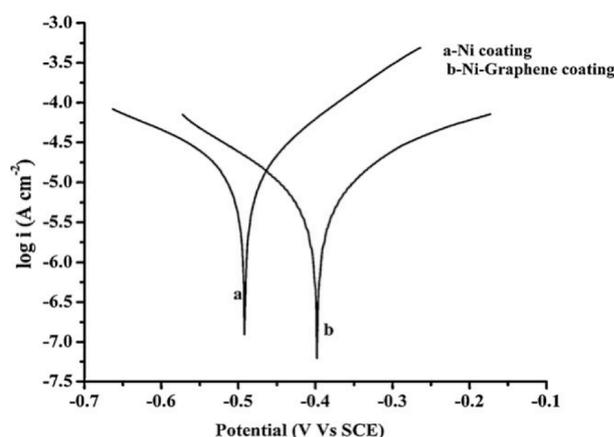
Figure 12: Tafel curves of Ni and Ni–Gr materials [186] (by courtesy of the publisher, Elsevier)

A more exotic approach is to develop metal-Gr composites for corrosion protection, although this strategy may have more disadvantages than advantages like with the metal-CNTs [179-182] because of the CNT activity towards catalytic oxygen reduction reactions [183-185]. Electrodeposited thin films of nickel matrix with Gr inclusion indicated a slight improvement in general protection performance; ennobled OCPs by ~0.1 V and steady corrosion currents reduced by over 50% were observed in comparison with pure nickel coated mild steel [186]. As a result of greatly affected nucleation and growth rate, measurement results are much more related to the highly altered crystalline structure of the nickel host matrix (lower electrical conductivity was unfortunately not characterised) as a consequence of the inclusion of Gr sheets, but differences in the texture and hardness of pure nickel and the Gr-nickel coatings were reported. Similarly to many other works, protection efficacy was not evaluated and reported over long-term immersion testing. On the one hand, the Tafel analysis of Ni and Ni–Gr performed in 3.5 wt.% NaCl solution (in *Fig.12*) was carried out using high voltage rate scanning without ohmic drop compensation, leading to distorted current legs over both potential regions (from the open-circuit or free corrosion potential of the electrode). A less representative transient is due to the low proportion of Faradaic charge-transfer current primarily assigned to the credible corrosion rate assessment. Probably due to these reasons, there is not much difference between the given corrosion currents of Ni and Ni-Gr samples. The Gr incorporation, cathodic and anodic Tafel regions exhibited almost ideal symmetry. This should mean the same scale of activation (expressed as a transfer function) in the cathodic (reduction and deposition) and anodic (oxidation and dissolution and/or deposition) processes. This feature is a kind of ineffectively blocked electrode behaviour, a quite reversible characteristic with less viable hindrance either in the anodic or cathodic processes. This suggestion seems to be confirmed by the reported similar corrosion currents.

On the other hand, the adequacy of EIS data may also be doubtful as basic requirements of system stability (in the time domain), linearity (immediate
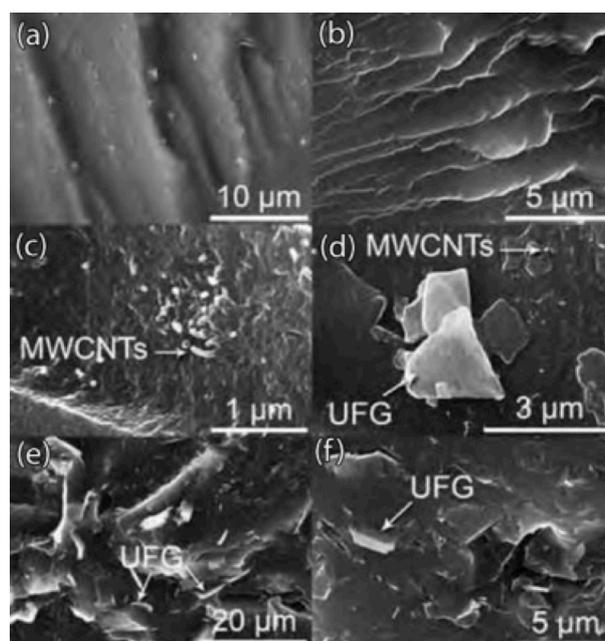
Figure 13: Cryo-fractured SEM images of (a) and (b), PEI, (c) and (d) 2 wt.% NFG/MWCNT/PEI, (e) and (f) 20 wt.% UNFG/PEI [187] (published with permission from the American Ceramic Society Bulletin)

response), and causality (the unquestionable relationship between perturbation and type-degree of response) might not be met. The time frame of 5 min to wait for a stationary or a steady-state condition then making measurements is in itself questionable. The large variation in the solution resistance (between the pure and composite Ni coating) is surely inadequate. In addition, the highly varied interfacial and coating resistance (swop between them) along with a hundred times difference in the coating capacitance are also unbelievable and can lead only to false interpretation. The depressed semicircle is a result of the complexity of the Ni/Gr electrode interface. The electrode was a porous electron and ion conductor and such systems need to be carefully handled (modelled with at least two transmission lines) when evaluation of impedance spectra in relation to structure and mechanistic behaviour of a complicated interface comes into the picture. Therefore, presented data and the corresponding evaluation are not appropriate and even misleading.

An innovative approach is the composite of multi-walled CNTs (MWCNTs) and exfoliated Gr embedded in polyether-imide (PEI) serving as a coherent and pore-free matrix [187]. This endeavour was made in an attempt to achieve a uniform dispersion of the nano-size additives in the host matrix (without functionalisation of the filler phase) and good adhesion of tailored PEI to the substrate, facilitating the maximum attainable corrosion protection performance. The protection mechanism was delineated according to the following: the composite protects steel alloys by a complex, active–passive way, serving as a physical barrier to water permeation against the evolution of ion channels to the metal surface and passivating metal surface by electron depletion at the interface based on the Schottky barrier effect [188-190].

Exfoliated but non-functionalised Gr was derived from ultrasonication in a solution of N-methyl-pyrrolidone (NMP). PEI as a host matrix was selected for its flexibility under a high glass transition temperature (155 ºC), thermal stability, firm resistance against radiation and appropriate interaction with carbon fillers *via* π–π stacking [191-193]. *In situ* synthesis and the imidisation protocol yielded coatings with appropriate adhesion properties to the steel surface of a thickness of 15–20 μm, whereas agglomeration and phase segregation of the fillers were low with the MWCNTs.

The SEM investigation revealed a fairly smooth surface finish without cracks or visible pinholes in the coatings. No phase segregation of the Gr and MWCNT fillers was observed in the mixtures around the interface as its clearly seen in *Fig.13* of the cryo-fractured samples of non-functionalised Gr (NFG), carbon nanotubes and PEI in 2 wt.% (c and d) and NFG/PEI of 20 wt.% (e, f) in comparison with pure PEI (a and b). OCPs were positively shifted by ~0.6 V, meaning a remarkable ennoblement. This was inferred, although substrates with high resistance coatings showing notable positive potentials must be at least partially attributed to the potential drop through the coatings [194] as a result of masking the real corrosion potential of the metal-coating interface to various extents. A potentiodynamic test indicated current densities of several orders of magnitudes lower ($\sim10^{-9}$ A cm^{-2}) for composite coated specimens compared to bare low-alloy and galvanised steel samples with current densities of $\sim10^{-5}$ A cm^{-2}. Immersion tested samples were characterised quantitatively with regards to the perspective of corrosion rate or protection of steel substrates by weight-loss measurement (according to the ASTM-G1 protocol) after 3,144 hours. The substrate coated with 20 wt% Gr-PEI performed best with ca. thousand times lower corrosion rates (ca. 8.5×10^{-4} mm year^{-1}) than bare low-alloy steel samples (ca. 1.2×10^{-1} mm year^{-1}). However, the PEI film containing Gr and MWCNT (ca. 9.2×10^{-3} mm year^{-1}) manifested a moderate average corrosion rate of ca. 5.5×10^{-4} mm year^{-1} compared to the neat PEI film.

Concluding Remarks

Overall, it is evident that atomic pure graphene films lack uniformity and partly this leads to insufficient instantaneous protection performance. Therefore, some sort of physical and/or chemical post-treatment is needed to facilitate long-term protection performance

From direct deposition techniques, the broadly and thoroughly investigated CVD technique is undoubtedly a viable alternative when large-scale continuous surface finishing is highly demanded. However, it is far from the best choice when corrosion protection efficiency and durability are the priorities, due to the complexity of growing cohesive films on the surface of metal substrates with preferable orientation of the grains, and weak adherence of Gr films sensitive to mechanical and surface charging effects.

When performance is an essential preference then spin-coating, kinetic spray, electro-spinning, and electrophoretic deposition should be employed depending on the complexity of the surface, conductivity of the substrate and additional technological aspects in relation to further processes. Among them, spin-coating and kinetic spray techniques develop good quality and homogeneous coatings of greatly varied thicknesses. Thin and ultra-thin films feature high a dispersity, and an even distribution and sometimes even improved quality of the nanoparticles. When thicker coatings are obtained on electrically conducting substrates with strong adhesion to complex surface geometries without the matter of a certain degree of alteration of the nanoparticles, then electrophoretic deposition must be used.

With respect to the utilisation of pure ultra-thin Gr films and thicker coatings as Gr-filled composites, there are ways to achieve further progress in relation to the ultimate aim of improved corrosion protection. For example, thin Gr films can still be used as a primary metal surface finishing. Alternatively, strongly adhering matrices with Gr filler might obviously mean a far more robust solution with protection efficacy targeting high performance demands for real situations. Parameters like the combination and concentration of the fillers, and compatibility with the host matrices besides matching complementary technological requirements need to be carefully optimised to gain appropriate barrier and/or inhibition functions.

Acknowledgement

REFERENCES

[1] NOVOSELOV K.S., GEIM A.K., MOROZOV S.V., JIANG D., ZHANG Y., DUBONOS S.V., GRIGORIEVA I.V., FIRSOV A.A.: Electric Field Effect in Atomically Thin Carbon Films, Science, *2004*, 306(5696), 666–669

[2] NOVOSELOV K.S., GEIM A.K., MOROZOV S.V., JIANG D., KATSNELSON M.I., GRIGORIEVA I.V., DUBONOS S.V., FIRSOV A.A.: Two-dimensional gas of massless Dirac fermions in graphene, Nature, *2005*, 438(7065), 197–200

[3] BALANDIN A.A.: Thermal properties of graphene and nanostructured carbon materials. Nat. Mater., *2011*, 10(8), 569–581

[4] RAFIEE J., MI X., GULLAPALLI H., THOMAS A.V., YAVARI F., SHI Y., AJAYAN P.M., KORATKAR N.A.: Wetting transparency of graphene, Nat. Mater., *2012*, 11(3), 217–222

[5] RAFIEE J., RAFIEE M.A., YU Z.-Z., KORATKAR N.: Superhydrophobic to superhydrophilic wetting control in graphene films, Adv. Mater., *2010*, 22(19), 2151– 2154

[6] SUTTER P., MINNITI M., ALBRECHT P., FARÍAS D., MIRANDA R., SUTTER E.: A high-reflectivity, ambient-stable graphene mirror for neutral atomic and molecular beams, Appl. Phys. Lett., *2011*, 99(21), Art. No. 211907

[7] BUNCH J.S., VERBRIDGE S.S., ALDEN J.S., VAN DER ZANDE A.M, PARPIA J.M., CRAIGHEAD H.G., MCEUEN P.L.: Impermeable atomic membranes from graphene sheets, *2008*, Nano Lett., 8(8), 2458–2462

[8] NAIR R.R., WU H.A., JAYARAM P.N., GRIGORIEVA I.V., GEIM A.K.: Unimpeded permeation of water through helium-leak-tight graphene-based membranes, Science, *2012*, 335(6067), 442–444

[9] LEENAERTS O., PARTOENS B., PEETERS F.M.: Graphene: a perfect nanoballoon, Appl. Phys. Lett., *2008*, 93(19), Art. No. 193107

[10] BERRY V.: Impermeability of graphene and its applications, Carbon, *2013*, 62(1), 1–10

[11] CHEN S., BROWN L., LEVENDORF M., CAI W., JU S.-Y., EDGEWORTH J., LI X., MAGNUSON C.W., VELAMAKANNI A., PINER R.D., KANG J., PARK J., RUOFF R.S.: Oxidation resistance of graphene-coated Cu and Cu/Ni alloy, ACS Nano, *2011*, 5(2), 1321–1327

[12] LIU L., RYU S.M., TOMASIK M.R., STOLYAROVA E., JUNG N., HYBERTSEN M.S., STEIGERWALD M.L., BRUS L.E., FLYNN G.W.: Graphene oxidation: thickness-dependent etching and strong chemical doping, Nano Lett., *2008*, 8(7) 1965–1970

[13] SURWADE S.P., LI Z.T., LIU H.T.: Thermal oxidation and unwrinkling of chemical vapor deposition-grown graphene, J. Phys. Chem. C, *2012*, 116(38) 20600–20606

[14] PRASAI, D. TUBERQUIA J.C., HARL R.R., JENNINGS G.K., ROGERS B.R., BOLOTIN K.I.: Graphene: corrosion-inhibiting coating, ACS Nano, *2012*, 6(2), 1102–1108

[15] DE HEER W.A., BERGER C., WU X., FIRST P.N., CONRAD E.H., LI X., LI T., SPRINKLE M., HASS J., SADOWSKI M.L., POTEMSKI M., MARTINEZ G.: Epitaxial graphene, Solid State Comm., *2007*, 143(1-2) 92–100

[16] SHIVARAMAN S., CHANDRASHEKHAR M.V.S., BOECKL J., SPENCER M.G.: Thickness estimation of epitaxial graphene on SiC using attenuation of substrate Raman intensity, J. Electron. Mater., *2009*, 38(6), 725–730

[17] RATINAC K.R., YANG W., GOODING J.J., THORDARSON P., BRAET F.: Graphene and related materials in electrochemical sensing, Electroanalysis, 2011, 23(8)3, 803–826

[18] BROWNSON D.A.C., MUNRO L.J., KAMPOURIS D.K., BANKS C.E.: Electrochemistry of graphene: not such a beneficial electrode material? RSC Adv., *2011*, 1(6), 978–988

[19] RAMAN R.K.S., TIWARI A.: Graphene: The thinnest known coating for corrosion protection, J. Org. Materials, *2014*, 66(4), 637–642

[20] TONG Y., BOHM S., SONG M.: Graphene based materials and their composites as coatings, Austin J. Nanomed. Nanotechnol., *2013*, 1(1), 1–16

[21] REINA A., JIA X., HO J., NEZICH D., SON H., BULOVIC V., DRESSELHAUS M.S., KONG J.: Large area, few-layer graphene films on arbitrary substrates by chemical vapor deposition, Nano Lett., *2009*, 9(1) 30–35

[22] KIM K.S., ZHAO Y., JANG H., LEE S.Y., KIM J.M., KIM K.S., AHN J.-H., KIM P., CHOI J.-Y., HONG B.H.: Large-scale pattern growth of graphene films for stretchable transparent electrodes, Nature, *2009*, 457(7230), 706–710

[23] CORAUX J., N'DIAYE A.T., BUSSE C., MICHELY T.: Structural coherency of graphene on Ir(111), Nano Lett., 2008, 8(2), 565–570

[24] SUTTER P.W., FLEGE J.-I., SUTTER E.A.: Epitaxial graphene on ruthenium, Nat. Mater., *2008*, 7(5), 406–410

[25] LI X., CAI W., AN., KIM S., NAH J., YANG D., PINER R., VELAMAKANNI A., JUNG I., TUTUC E., BANERJEE S.K., COLOMBO L., RUOFF R.S.: Large-area synthesis of high-quality and uniform graphene films on copper foils, Science, *2009*, 324(5932) 1312–1314

[26] LEE Y., BAE S., JANG H., JANG S., ZHU S.E., SIM S.H., SONG Y.I., HONG B.H., AHN J.H.: Wafer-scale synthesis and transfer of graphene films, Nano Lett., *2010*, 10(2), 490–493

[27] LI X.S., ZHU Y.W., CAI W.W., BORYSIAK M., HAN B., CHEN D., PINER R.D., COLOMBO L., RUOFF R.S.: Transfer of large-area graphene films for high-performance transparent conductive electrodes, Nano Lett., *2009*, 9(12), 4359–4363

[28] PARK H.J., MEYER J., ROTH S., SKÁKALOVA V.: Growth and properties of few-layer graphene prepared by chemical vapor deposition, Carbon, *2010*, 48(4), 1088–1094

[29] BAE S., KIM H., LEE Y., XU X., PARK J.-S., ZHENG Y., BALAKRISHNAN J., LEI T., KIM H.R., SONG Y.I., KIM Y.-J., KIM K.S., ÖZYILMAZ B., AHN J.-H., HONG B.H., IIJIMA S.: Roll-to-roll production of 30-inch graphene films for transparent electrodes, Nat. Nanotech., *2010*, 5(8), 574–578

[30] LI X., MAGNUSON C.W., VENUGOPAL A., AN J., SUK J.W., HAN B., BORYSIAK M., CAI W., VELAMAKANNI A., ZHU Y., FU L., VOGEL E.M., VOELKL E., COLOMBO L., RUOFF R.S.: Graphene films with large domain size by a two-step chemical vapor deposition process, Nano Lett., *2010*, 10(1), 4328– 4334

[31] LI X., MAGNUSON C.W., VENUGOPAL A., TROMP R.M., HANNON J.B., VOGEL E.M., COLOMBO L., RUOFF R.S.: Large-area graphene single crystals grown by low-pressure chemical vapor deposition of methane on copper, J. Am. Chem. Soc., *2011*, 133(9), 2816–2819

[32] VLASSIOUK I., SMIRNOV S., SURWADE S.P., REGMI M., SRIVASTAVA N., FEENSTRA R., ERES GY., PARISH C., LAVRIK N., DATSKOS P., DAI S., FULVIO P.: Graphene nucleation density on copper: fundamental role of background pressure, J. Phys. Chem. C, *2013*, 117(37), 18919–18926

[33] BERGER C., SONG Z.M., LI X.B., WU X.S., BROWN N., NAUD C., MAYO D., LI T.B., HASS J., MARCHENKOV A.N, CONRAD E.H., FIRST P.N., DE HEER W.A.: Electronic confinement and coherence in patterned epitaxial graphene, Science, *2006*, 312(5777), 1191–1196

[34] BERGER C., SONG Z.M., LI T.B., LI X.B., OGBAZGHI A.Y., FENG R., DAI Z.T., MARCHENKOV A.N., CONRAD E.H., FIRST P.N., DE HEER W.A.: Ultrathin epitaxial graphite: 2D electron gas properties and a route toward graphene-based nanoelectronics, J. Phys. Chem. B, *2004*, 108(52), 19912–19916

[35] VLASSIOUK I., FULVIO P., MEYER H., LAVRIK N., DAI S., DATSKOS P., SMIRNOV S.: Large-scale atmospheric pressure chemical vapor deposition of graphene, Carbon, *2013*, 54(4), 58–67

[36] VAN GILS S., LE PEN C., HUBIN A., TERRYN H., STIJNSB E.: Electropolishing of copper in H₃PO₄ *ex situ* and *in situ* optical characterization, J. Electrochem. Soc., *2007*, 154(3), C175–C180

[37] SHIVAREDDY S., BAE S., BRANKOVIC S.: Cu surface morphology evolution during electropolishing, Electrochem. Solid State Lett., *2008*, 11(1), D13–D17

[38] WOOD J.D., SCHMUCKER S.W., LYONS A.S., POP E., LYDING J.W.: Effects of polycrystalline Cu substrate on graphene growth by chemical vapor deposition, Nano Lett., *2011*, 11(1), 4547–4554

[39] VLASSIOUK I., REGMI M., FULVIO P., DAI S., DATSKOS P., ERES G., SMIRNOV S.: Role of hydrogen in chemical vapor deposition growth of large single-crystal graphene, ACS Nano, *2011*, 5 (7), 6069–6076

[40] FLORES C.B., LÓPEZ D.M.: Multilayer graphene synthesized by CVD using liquid hexane as the carbon precursor, World J. Cond. Matter Phys., *2011*, 1(4), 157–160

[41] KIRKLAND N.T., SCHILLER T., MEDHEKAR N., BIRBILIS N.: Exploring graphene as a corrosion protection barrier, Corr. Sci., *2012*, 56(1), 1–4

[42] KOUSALYA A.S., KUMAR A., PAUL R., ZEMLYANOV D., FISHER T.S.: Graphene: An effective oxidation barrier coating for liquid and two-phase cooling systems, Corr. Sci., *2013*, 69(1), 5–10

[43] KALITA G., AYHAN M.E., SHARMA S., SHINDE S.M., GHIMIRE D., WAKITA K., UMENO M., TANEMURA M.: Low temperature deposited graphene by surface wave plasma CVD as effective oxidation resistive barrier, Corr. Sci., *2014*, 78(1), 183–187

[44] FRIEDRICH H., JACOBY G., MEISTER C.G.: Quantum reflection by Casimir–van der Waals potential tails, Phys. Rev. A, *2002*, 65(3), 032902–032915

[45] SUTTER E., ALBRECHT P., CAMINO F.E., SUTTER P.: Monolayer graphene as an ultimate chemical passivation layer for arbitrarily shaped metal surfaces, Carbon, *2010*, 48(15), 4414–4420

[46] SINGH RAMAN R.K., BANERJEE P.C., LOBO D.E., GULLAPALLI H., SUMANDASA M., KUMAR A., CHOUDHARY L., TKACZ R., AJAYAN P.M., MAJUMDER M.: Protecting copper from electrochemical degradation by graphene coating, Carbon, *2012*, 50(11), 4040–4045

[47] HSIEH Y.-P., HOFMANN M., CHANG K.-W., JHU J.G., LI Y.-Y., YAO CHEN K., YANG C.C., CHANG W.-S., CHEN L.-C.: Complete corrosion inhibition through graphene defect passivation, ACS Nano, *2014*, 8(1), 443–448

[48] HUH J.-H., KIM S.H., CHU J.H., KIM S.Y., KIM J.H., KWON S.-Y.: Enhancement of seawater corrosion resistance in copper using an acetone-derived graphene coating, Nanoscale, *2014*, 6(8), 4379–4386

[49] NAYAK P.K., HSU C.-J., WANG S.-C., J.C., SUNG HUANG J.-L.: Graphene coated Ni films: A protective coating, Thin Solid Films, *2013*, 529(1), 312–316

[50] XIA J., CHEN F., LI J., TAO N.: Measurement of the quantum capacitance of graphene, Nat. Nanotechnol., 2009, 4(8), 505–509

[51] WANG S., ZHANG L., XIA Z., ROY A., CHANG D.W., BAEK J.-B., DAI L.: BCN graphene as an efficient metal-free electrocatalyst for the oxygen reduction reaction, Angew. Chem. Int. Ed., *2012*, 51(17), 4209–4212

[52] ERNEST Y.: Dioxygen electrocatalysis: mechanism in relation to catalyst structure, J. Mol. Catal., *1986*, 38(1-2), 5–25

[53] SREEVATSA S., BANERJEE A., HAIM G.: Graphene as a permeable ionic barrier, ECS Trans., *2009*, 19(5), 259–264

[54] ROBINSON Z.R., TYAGI P., MURRAY T.M., VENTRICE C.A., CHEN S., MUNSON A., MAGNUSON C.W., RUOFF R.S.: Substrate grain size and orientation of Cu and Cu–Ni foils used for the growth of graphene films, J. Vac. Sci. Technol. A, Vac. Surf. Films, *2011*, 30(1), Art. No. 011401

[55] LI X., CAI W., COLOMBO L., RUOFF R.S.: Evolution of graphene growth on Ni and Cu by carbon isotope labeling, Nano Lett., *2009*, 9(12), 4268–4272

[56] SRIVASTAVA A., GALANDE C., CI L., SONG L., RAI C., JARIWALA D., KELLY K.F., AJAYAN P.M.: Multilayer graphene synthesized by CVD using liquid hexane as the carbon precursor, Chem. Mater., *2010*, 22(11), 3457–3461

[57] MIAO C., ZHENG C., LIANG O., XIE Y.-H.: Physics and applications of graphene - experiments, Ed. MIKHAILOV S., InTech Europe, Rijeka, Croatia, *2011*

[58] ZHOU X.: Graphene oxidation barrier coating, Ph.D. Thesis, University of Colorado, Boulder, CO USA, *2011*

[59] SCHRIVER M., REGAN W., GANNETT W.J., ZANIEWSKI A.M., CROMMIE M.F., ZETTL A.: Graphene as a long-term metal oxidation barrier: worse than nothing, ACS Nano, *2013*, 7(7), 5763–5768

[60] HADDAD F.S., THAKRAR R.R., HART A.J., SKINNER J.A., NARGOL A.V.F., NOLAN J.F., GILL H.S., MURRAY D.W., BLOM A.W., CASE C.P.: Metal-on-metal bearings, J. Bone Joint Surg. Br., *2011*, 93(B), 572–579

[61] KOROVESSIS P., PETSINIS G., REPANTI M., REPANTIS T.: Metallosis: After contemporary metal-on-metal total hip arthroplasty five to nine-year follow-up, J. Bone Joint Surg. Am., *2006*, 88(6), 1183–1191

[62] KAW M., SINGH S., GAGNEJA H., AZAD P.: Role of self-expandable metal stents in the palliation of malignant duodenal obstruction, Surg. Endosc., *2003*, 17(4), 646–650

[63] BARON T.H.: Expandable metal stents for the treatment of cancerous obstruction of the gastrointestinal tract, New Engl. J. Med., *2001*, 344(22), 1681–1687

[64] CORTIZO M., DEMELE M., CORTIZO A.: Metallic dental material biocompatibility in osteoblastlike cells, Biol. Trace Elem. Res., *2004*, 100(2), 151–168

[65] CRAIG R.G., POWERS J.M. (Eds.): Restorative dental materials, 11th Ed., Mosby Inc., St Louis, MO, USA, *2010*

[66] TONDATO F., NG D.W., SRIVATHSAN K., ALTEMOSE G.T., HALYARD M.Y., SCOTT L.R.: Radiotherapy-induced pacemaker and implantable cardioverter defibrillator malfunction, Expert Rev. Med. Dev., *2009*, 6(3), 243–249

[67] ISHII K., KODANI E., MIYAMOTO S., OTSUKA T., HOSONE M., OGATA K., SATO W., MATSUMOTO S., TADERA T. IBUKI KUSAMA Y., ATARASHI H.: Pacemaker Contact Dermatitis: The effective use of a polytetrafluoroethylene sheet, Pacing Clin. Electrophysiol., *2006*, 29(11), 1299–1302

[68] SCHMIDT C., IGNATIUS A.A., CLAES L.E.: Proliferation and differentiation parameters of human osteoblasts on titanium and steel surfaces, J. Biomed. Mater. Res., *2001*, 54(2), 209–215

[69] KURTZ S.M., MURATOGLU O.K., EVANS M., EDIDIN A.A.: Advances in the processing, sterilization, and crosslinking of ultra-high molecular weight polyethylene for total joint arthroplasty, Biomaterials, *1999*, 20(18) 1659–1688

[70] DAVIS J.R.: Handbook of materials for medical devices, First Ed., ASM International, Materials Park, OH, USA, *2003*

[71] ATWOOD R.C., LEE P.D., CURTIS R.V.: Modeling the surface contamination of dental titanium investment castings, Dent. Mater., *2005*, 21(2), 178–186

[72] CAICEDO M.S., PENNEKAMP P.H., MCALLISTER K., JACOBS J.J., HALLAB N.J.: Soluble ions more than particulate cobalt-alloy implant debris induce monocyte costimulatory molecule expression and release of proin flammatory cytokines critical to metal-induced lymphocyte reactivity, J. Biomed. Mater. Res., *2010*, A93A(4), 1312–1321

[73] ROMESBURG J.W., WASSERMAN P.L., SCHOPPE C.H.: Metallosis and Metal-Induced Synovitis Following Total Knee Arthroplasty, Rev. Radiograp. CT Findings, *2010*, 4(9), 7–17

[74] JACOBS J.J., HALLAB N.J.: Loosening and osteolysis associated with metal-on-metal bearings: A local effect of metal hypersensitivity?, J. Bone Joint Surg. Am., *2006*, 88(6), 1171–1172

[75] YANG L., SHELDON B.W., WEBSTER T.J.: Orthopedic nano diamond coatings: Control of surface properties and their impact on osteoblast adhesion and proliferation, J. Biomed. Mater. Res., *2009*, A91A(2), 548–556

[76] PECHEVA E., PRAMATAROVA L., FINGAROVA D., HIKOV T., DINEVA I., KARAGYOZOVA Z., STAVREV S.: Advanced materials for metal implant coatings, J. Optoelectron. Adv. Mater., *2009*, 11(9), 1323–1326

[77] PODILA R., MOORE T., ALEXIS F., RAO A.M.: Graphene coatings for enhanced hemo-compatibility of nitinol stents, Rev. Soc. Chem. Adv., *2013*, 3(1), 1660–1665

[78] PODILA R., MOORE T., ALEXIS F.R.: Graphene coatings for biomedical implants, J. Vis. Exp., *2013*, 73(1), Art. No. e50276

[79] ZHANG W., LEE S., MCNEAR K.L., CHUNG T.F., LEE S., LEE K., CRIST S.A., RATLIFF T.L., ZHONG Z. YANG C.: Use of graphene as a protection film in biological environments, Sci. Rep. *2014*, 4(4), 4097

[80] NEUHAUSER E.F., LOEHR R.C., MILLIGAN D.L., MALECKI M.R.: Toxicity of metals to the earthworm *Eisenia fetida*, Biol. Fert. Soils, *1985*, 1(3), 149–152

[81] KOENIG S.P., BODDETI N.G., DUNN M.L., BUNCH J.S.: Ultrastrong adhesion of graphene membranes, Nat. Nano., 2011, 6(9), 543–546

[82] CAPRIOLI F., DECKER F., MARRANI A.G., BECCARI M., CASTRO V.D.: Copper protection by self-assembled monolayers of aromatic thiols in alkaline solutions, Phys. Chem. Chem. Phys., *2010*, 12(32), 9230–9238

[83] LUSK A.T., JENNINGS G.K.: Characterization of self-assembled monolayers formed from sodium s-alkyl thiosulfates on copper, Langmuir, *2001*, 17(25), 7830–7836

[84] BRINKER C.J., HURD A.J., SHUNK P.R., FRYE G.C., ASHLEY C.S.: Review of sol-gel thin film formation, J. Non-Cryst. Solids, *1992*, 147-148(1), 424–436

[85] SCRIVEN L.E.: in Better Ceramics through Chemistry III, Eds., BRINKER C.J., CLARK D.E., ULRICH D.R., The Materials Research Society, Pittsburgh, PA USA, *1988*, p.717

[86] LANDAU L.D., LEVICH B.G.: Dragging of a liquid by a moving plate, Acta Physiochim. U.R.S.S., *1942*, 17, 42–54

[87] DISLICH H.: Sol-gel: science, processes and products, J. Non-Cryst. Solids, *1997*, 80(1-3), 115–121

[88] ARFSTEN N.J., ERBLE A., OTTO J., REICH A.: Investigations on the angle-dependent dip coating technique, ADDC, for the for the Production of Optical Filters, J. Sol-Gel Sci. Techol., *1997*, 8(1-3), 1099–1104

[89] SCHMIDT H., in Chemistry, Spectroscopy and Application of sol-gel glasses, Ed., REISFELD R. Springer-Verlag, Berlin, Germany, *1992*, pp.119–152

[90] MEYERHOFER D.: Characteristics of resist films produced by spinning, J. Appl. Phys., *1978*, 49(7), 3993–3997

[91] LAI J.H.: An investigation of spin coating of electron resists, Polym. Eng. Sci., *1979*, 19(15), 1117–1121

[92] CHEN B.T.: Investigation of the solvent-evaporation effect on spin coating of thin films, Polym. Eng. Sci., *1983*, 23(7), 399–403

[93] VAN BOMMEL J.: Optical coatings for computer monitors and TV screens, Glass Res., *1997*, 10, 7

[94] O'BRIEN S.B.G., SCHWARTZ L.W.: Theory and modeling of thin film flows, Encyclopedia of Surface and Colloid Science, Ed. HUBBARD A., Marcel Dekker, New York, *2002*, pp. 5283–5297

[95] EVANS P.L., SCHWARTZ L.W., ROY R.V.: Three-dimensional solutions for coating flow on a rotating horizontal cylinder: Theory and experiment, Phys. Fluids, *2005*, 17(7), Art. No. 072102

[96] FLOCH H.G., BELLEVILLE G., PRIOTTON J.J.: Sol-Gel Optical Coatings For Lasers., Am. Ceram. Soc. Bull., *1995*, 74(10), 60–63

[97] FLOCH H.G., BELLEVILLE P.F.: Scratch-Resistant Single-Layer Antireflective Coating By A Low Temperature Sol-Gel Route, Proc. Spie, *1992*, 1758, 135–149

[98] MITSOULIS E., ATHANASOPOULOS G.: Numerical Simulation Of Blade-Over-Roll Coating Forming Flows, Comp. Met. Mater., *2010*, 10(4), 214–224

[99] NODA K., YAMAGAMI YAMAMOTO M., M., SATO H., ITAYA H., OKAJIMA A., YAMADA M.: Thin Film Formation By Direct Reverse Roll-Coating On Plastic Web, Jpn. J. Appl. Phys., *2003*, 42(1), 5722–5725

[100] ALONSO S., REÂGLAT O., BERTRAND F., CHOPLIN L., TANGUY P.A.: Process viscosity in reverse roll coating, Trans. I. Chem. E, *2001*, 79(Part A), 128–136

[101] EMSLIE A.G., BONNER F.T., PECK C.G.: Flow of a viscous liquid on a rotating disk, J. Appl. Phys., *1958*, 29(5), 858–862

[102] ACRIVOS A., SHAH M.J., PETERSEN E.E.: On the flow of a non-Newtonian liquid on a rotating disk, J. Appl. Phys., *1960*, 31(6), 963–968

[103] KIM D.-Y., SINHA-RAY S., PARK J.-J., LEE J.-G., CHA Y.-H., BAE S.-H., AHN J.-H., JUNG Y.C., KIM S.M., YARIN A.L., YOON S.S.: Self-healing reduced graphene oxide films by supersonic kinetic spraying, Adv. Funct. Mater. *2014*, 24(31), 4986–4995

[104] WANG S.J., GENG Y., ZHENG Q., KIM J.-K.: Fabrication of highly conducting and transparent graphene films, Carbon, *2010*, 48(6), 1815–1823

[105] YOON S., KIM H.J., LEE C.: Fabrication of an automotive heat exchanger using a kinetic spraying process, Surf. Coat. Technol., *2007*, 201(1) 9524–9532

[106] DZHURINSKIY D., MAEVA E., LESHCHINSKY EV., MAEV, R.GR.: Corrosion Protection of Light Alloys Using Low Pressure Cold Spray, J. Therm. Spray Technol., *2012*, 21(2) 304–313

[107] HUSSAIN T., MCCARTNEY D.G., SHIPWAY P.H., MARROCCO T.: Corrosion behavior of cold sprayed titanium coatings and free standing deposits, J. Therm. Spray Technol., *2011*, 20(1-2), 260–274

[108] YARIN A.L., KOOMBHONGSE S., RENEKER D.H.: Taylor cone and jetting from liquid droplets in the electrospinning of nanofibers, J. Appl. Phys., *2001*, 90(9), 4836–4846

[109] SHIN Y.M., HOHMAN M.M., BRENNER M.P., RUTLEDGE G.C.: Experimental characterization of electrospinning: the electrically forced jet and instabilities, Polymer, *2001*, 42(25), 9955–9967

[110] MENCHACA C., MANOUN B., MARTÍNEZ-BARRERA G., CASTANO V.M., LÓPEZ-VALDIVIA H.: In situ high-temperature Raman study of crystalline nylon 6,12 fibers gamma-irradiated in an argon atmosphere, J. Phys. Chem. Solids, *2006*, 67 (9), 2111–2118

[111] CAMPOS C.M., PÉREZ C.G., CASTAÑEDA I., GARCÍA-SÁNCHEZ M.A., GUARDIÁN R., URUCHURTU J.: Nylon/graphene oxideelectrospun composite coating, Internat. J. Polym. Sci., *2013*, 1, 1-9

[112] KAYA C.: Al_2O_3-Y-TZP/Al_2O_3 functionally graded composites of tubular shape from nano-sols using double-step electrophoretic deposition, J. Eur. Ceram. Soc., *2003*, 23(10), 1655–1660

[113] PUT S., VLEUGELS J., VAN DER BIEST O.: Functionally graded WC–co materials produced by electrophoretic deposition, Scr. Mater., *2001*, 45(10), 1139–1145

[114] SARKAR P., DATTA S., NICHOLSON P.S.: Functionally graded ceramic/ceramic and metal/ceramic composites by electrophoretic deposition, Compos. B-Eng., *1997*, 28(1), 49–56

[115] LEE S.B., CHOI O., LEE W., YI J.W., KIM B.S., BYUN J.H., YOON M.K., FONG H., THOSTENSON E.T., CHOU T.W.: Processing and characterization of multi-scale hybrid composites reinforced with nanoscale carbon reinforcements and carbon fibers, Compos. Part A, *2001*, 42(4), 337–344

[116] ZHITOMIRSKY I.: Hydroxyapatite coatings and fibers, Mater. Lett., *2000*, 42, 262

[117] YOU C., JIANG D.L., JAN S.H.: SiC/TiC laminated structure shaped by electrophoretic deposition, Ceram. Int., *2004*, 30(5), 813–815

[118] VAN DE PERRE L., VAN DER BIEST O.: Composite SiC-graphite interlayers for crack deflection in ceramic laminates, Silic. Ind., *1998*, 63, 39–43

[119] KAYA C., KAYA F., SU B., THOMAS B., BOCCACCINI A.R.: Structural and functional thick ceramic coatings by electrophoretic deposition, Surf. Coat. Technol., *2005*, 191(1), 303–310

[120] YANAGIDA A., NAKAJIMA Y., KAMESHIMA N., YOSHIDA T., WATANABE K., OKADA K.: Preparation of a crack-free rough titania coating on a stainless steel mesh by electrophoretic deposition, Mater. Res. Bull., *2005*, 40(1), 1335–1344

[121] PEIRO A.M., BRILLA E., PERAL J., DOMENECH X., AYLLON J.A.: Electrochemically assisted deposition of titanium dioxide on aluminium cathodes, J.Mater.Chem., *2002*, 12(1), 2769–2773

[122] BOCCACCINI A.R., ROETHER J.A., THOMAS B.J.C., SHAFFER M.S.P., CHAVEZ E., STOLL E., MINAY E.J.: 'The electrophoretic deposition of inorganic nanoscaled materials, J. Ceram. Soc. Jpn., *2006*, 114(1325), 1–14

[123] BESRA L., LIU M.: A review on fundamentals and applications of electrophoretic deposition number , EPD , Prog. Mater. Sci., *2007*, 52, 1

[124] VU Q.-T., PAVLIK M., HEBESTREIT N., PFLEGER J., RAMMELT U., PLIETH W.: Electrophoretic deposition of nanocomposites formed from polythiophene and metal oxides, Electrochim. Acta, *2005*, 51(6), 1117–1124

[125] PECH-RODRÍGUEZ W.J., GONZÁLEZ-QUIJANO D., VARGAS-GUTIÉRREZ G., RODRÍGUEZ-VARELA F.J.: Deposition of Vulcan XC-72 coatings on stainless steel bipolar plates by reverse pulsed dc voltage electrophoretic deposition number, EPD, for fuel cell applications, ECS Trans., *2014*, 58(12), 33–39

[126] LIH E.T.Y., ZAID R.T.M., LING T.L., CHONG K.F.: Facile corrosion protection coating from graphene, Int. J. Chem. Engin. Appl., *2012*, 3(6), 453–455

[127] SINGH B.P., NAYAK S., NANDA K.K., JENA B.K., BHATTACHARJEE S., BESRA L.: The production of a corrosion resistant graphene reinforced composite coating on copper by electrophoretic deposition, Carbon, *2013*, 61(1), 47–56

[128] HE W., ZHU L., CHEN H., NAN H., LI W., LIU H., WANG Y.: Electrophoretic deposition of graphene oxide as a corrosion inhibitor for sintered NdFeB, Appl. Surf. Sci., *2013*, 279(1), 416–423

[129] KOLBE H.: Decomposition of valeric acid by an electric current, Annal. Chem. Pharm., *1848*, 64(1), 339–341 (in German)

[130] KOLBE H.: Studies of the electrolysis of organic compounds, Annal. Chem. Pharm., *1849*, 69(1), 257–294 (in German)

[131] SINGH B.P., JENA B.K., BHATTACHARJEE S., BESRA L.: Development of the oxidation and corrosion resistanr hydrophobic graphene oxide-polymer composite coating on copper, Surf. Coat. Technol., *2013*, 232(1), 475–481

[132] WANG M., DUONG L.D., OH J.-S., MAI N.T., KIM S., HONG S., WANG T.H., LEE Y.K., NAM J.-D.: Large-Area, conductive and flexible reduced graphene oxide number, RGO, membrane fabricated by electrophoretic deposition number, EPD, ACS Appl. Mater. Interfaces, *2014*, 6(3), 1747–1753

[133] MEYERS J.P., DARLING R.M.: Model of carbon corrosion in PEM fuel cells, J. Electrochem. Soc., *2006*, 153(8), A1432–A1442

[134] PRABAKAR S.J.R., HWANG Y.-H., BAE E.G., LEE D.K., PYO M.: Graphene oxide as a corrosion inhibitor for the aluminium current collector in lithium ion batteries, Carbon, *2013*, 52(1), 128–136

[135] LOGAN B.E., HAMELERS B., ROZENDAL R., SCHRÖDER U., KELLER J., FREGUIA S., AELTERMAN P., VERSTRAETE W., RABAEY K.: Microbial fuel cells: methodology and technology, Environ. Sci. Technol., *2006*, 40(17), 5181–5192

[136] KRISHNAMURTHY A., GADHAMSHETTY V., MUKHERJEE R., CHEN Z., REN W., CHENG H-M., KORATKAR N.: Passivation of microbial corrosion using a graphene coating, Carbon, *2013*, 56(1), 45–49

[137] SAHU S.C., SAMANTARA A.K., SETH M., PARWAIZ S., SINGH B.P., RATH P.C., JENA B.K.: A facile electrochemical approach for the development of highly corrosion protective coatings using graphene nanosheets, Electrochem. Commun., *2013*, 32(1), 22–26

[138] WESSLING B.: Passivation of metals by coating with polyaniline: Corrosion potential shift and morphological changes, Adv. Mater., *1994*, 6(3), 226–228

[139] DEBERRY D.W.: Modification of the electrochemical and corrosion behavior of stainless steels with an electroactive coating, J. Electrochem. Soc., *1985*, 132(5), 1022–1026

[140] WEI Y., WANG J., JIA X., YEH J.M., SPELLANE P.: Polyaniline as corrosion protection coatings on cold rolled steel, Polymer, *1995*, 36(23), 4535–4537

[141] YEH J.M., LIOU S.J., LAI C.Y., WU P.C., TSAI T.Y.: Enhancement of the corrosion protection effect in polyaniline *via* the formation of polyaniline-clay nanocomposite materials, Chem. Mater., *2001*, 13(3), 1131–1136

[142] IONITA M., PRUNA A.: Polypyrrole/carbon nanotube composites: Molecular modeling and experimental investigation as an anti-corrosive coating, Prog. Org. Coat., *2011*, 72(4), 647–652

[143] MARTINA V., DE RICCARDIS M.F., CARBONE D., ROTOLO P., BOZZINI B., MELE C.: Electrodeposition of polyaniline–carbon nanotubes composite films and investigation on their role in corrosion protection of austenitic stainless steel by SNIFTIR analysis: J. Nanopart. Res., *2011*, 13(11), 6035–6047

[144] SALAM M.A., AL-JUAID S.S., QUSTI A.H., HERMAS A.A.: Electrochemical deposition of a carbon nanotube-poly o-phenylenediamine, composite on a stainless steel surface, Synth. Met., *2011*, 161(1-2), 153–157

[145] HU S.T., KONG X.H., YANG H., SUN R.H.: Anticorrosive films prepared by incorporating permanganate modified carbon nanotubes into the waterborne polyurethane polymer, Adv. Mater. Res., *2011*, 189–193, 1157–1162

[146] DE RICCARDIS M.F., MARTINA V., CARBONE D., ROTOLO P.: Functional characterisations of hybrid nanocomposite films based on polyaniline and carbon nanotubes, Adv. Sci. Technol., *2013*, 79(1), 81–86

[147] GERGELY A., PÁSZTI Z., MIHÁLY J., DROTÁR E., TÖRÖK T.: Galvanic corrosion prevention of the zinc-rich hybrid coatings facilitated by the percolating structure of the carbon nanotubes Part II: Protection properties and mechanism of the hybrid coatings, Prog. Org. Coat., *2014*, 77(2), 412–424

[148] GERGELY A., PÁSZTI Z., HAKKEL O., MIHÁLY J., KÁLMÁN E.: Corrosion protection of cold-rolled steel with alkyd paint coatings composited with various microstructure arranged polypyrrole-modified nano-size alumina and carbon nanotubes, Mater. Sci. Eng. B, *2012*, 177(18), 1571–1682

[149] GERGELY A., PÁSZTI Z., HAKKEL O., BERTÓTI I., MIHÁLY J., TÖRÖK T.: Investigation of modified carbon nanotubes/aluminium-oxide monohydrate containing zinc-rich hybrid paint coatings, Korr. Figy., *2013*, 53(1), 3–24 (in Hungarian)

[150] GERGELY A., TÖRÖK T.: Optimally balanced active-passive corrosion protection by zinc-rich paint coatings featuring proper hybrid formulation with polypyrrole modified carbon nanotubes, Mater. Sci. For., *2013*, 752(1), 275–283

[151] GERGELY A., PÁSZTI Z., BERTÓTI I., TÖRÖK T., MIHÁLY J., KÁLMÁN E.: Novel zinc-rich epoxy paint coatings with hydrated alumina and carbon nanotubes supported polypyrrole for corrosion protection of low carbon steel Part II: Corrosion prevention behaviour of the hybrid paint coatings, Mater. Corr., *2013*, 64(12), 1091–1103

[152] GERGELY A., TÖRÖK T., PÁSZTI Z., BERTÓTI I., MIHÁLY J., KÁLMÁN E.: Zinc-rich paint coatings containing either ionic surfactant-modified or functionalized multi-walled carbon nanotube-supported polypyrrole utilized to protect cold-rolled steel against corrosion, Nova Science Publishers, Inc., Part II: Application of Carbon nanotubes, *2012*, Chapter 10th, pp. 211–258.

[153] JEON H., PARK J., SHON M.: Corrosion protection by epoxy coating containing multi-walled carbon nanotubes, J. Ind. Engin. Chem., *2013*, 19(3), 849–853

[154] SHOW Y., NAKASHIMA T., FUKAMI Y.: Anticorrosion coating of carbon nanotube/polytetra-fluoroethylene composite film on the stainless steel bipolar plate for proton exchange membrane fuel cells, J. Nanomater., *2013*, Article ID 378752

[155] DESHPANDE P.P., VATHARE S.S., VAGGE S.T., TOMŠÍK E., STEJSKAL J.: Conducting polyaniline/multi-wall carbon nanotubes composite paints on low carbon steel for corrosion protection: Electro-chemical investigations, Chem. Pap., *2013*, 67(8), 1072–1078

[156] SREEVATSA S.: Carbon nanotube electronic structures as anti-corrosion coatings, PhD Thesis, State University of New Jersey, Newark, USA, 2009

[157] PRAVEEN B.M., VENKATESHA T.V., ARTHOBA NAIK Y., PRASHANTHA K.: Corrosion studies of carbon nanotubes–Zn composite coating, Surf. Coat. Technol., *2007*, 201(12), 5836–5842

[158] CHEN X.H., CHEN C.S., XIAO H.N., CHENG F.Q., ZHANG G., YI G.J.: Corrosion behavior of carbon nanotubes–Ni composite coating, Surf. Coat. Technol., *2005*, 191(2-3), 351–356

[159] MONTEMOR M.F., FERREIRA M.G.S.: Analytical characterisation and corrosion behaviour of bis-aminosilane coatings modified with carbon nanotubes activated with rare-earth salts applied on AZ31 Magnesium alloy, Surf. Coat. Technol., *2008*, 202(19), 4766–4774

[160] RAFIEE M.A., RAFIEE J., WANG Z., SONG H., YU Z.Z., KORATKAR N.: Enhanced mechanical properties of nanocomposites of low graphene content, ACS Nano, *2009*, 3(12), 3884–3890

[161] ZHANG W.L., LIU Y.D., CHOI H.J.: Fabrication of semiconducting graphene oxide/polyaniline composite particles and their electrorheological response under an applied electric field, Carbon, *2012*, 50(1), 290–296

[162] Zhang W.L., Park B.J., Choi H.J.: Colloidal graphene oxide/polyaniline nanocomposite and its electrorheology, Chem. Commun., *2010*, 46(30), 5596–5598

[163] Stankovich S., Piner R.D., Nguyen S.T., Ruoff R.S.: Synthesis and exfoliation of isocyanate-treated graphene oxide nanoplatelets, Carbon, *2006*, 44(15), 3342–3347

[164] Compton O.C., Kim S., Pierre C., Torkelson J.M., Nguyen S.T.: Crumpled graphene nanosheets as highly effective barrier property enhancers, Adv. Mater., *2010*, 22(42), 4759–4763

[165] Kim H.W., Miura Y., Macosko C.W.: Graphene/polyurethane composites for improved gas barrier and electrical conductivity, Chem. Mater., *2010*, 22(11), 3441–3450

[166] Kim H.W., Macosko C.W.: Dispersing organoclay in polystyrene melts: roles of stress and diffusion, Macromolecules, *2008*, 41(9), 3317–3327

[167] Kalaitzidou K., Fukushima H., Drzal L.T.: Multifunctional polypropylene composites produced by the incorporation of exfoliated graphite nanoplatelets, Carbon, *2007*, 45(7), 1446–1452

[168] Kim H.W., Abdala A.A., Macosko C.W.: Graphene/polymer composites, Macromolecules, *2010*, 43(16), 6515–6530

[169] Potts J.R., Dreyer D.R., Bielawski C.W., Ruoff R.S.: Graphene-based polymer composites, Polymer, *2011*, 52(1), 5–25

[170] Chang C.-H., Huang T.-C., Peng C.-W., Yeh T.-C., Lu H.-I., Hung W.-I., Weng C.-J., Yang T.-I., Yeh J.-M.: Novel anticorrosion coatings prepared from polyaniline/graphene composites, Carbon, *2012*, 50(14), 5044–5051

[171] Liu T., Yin Y., Chen S., Chang X., Cheng S.: Super-hydrophobic surfaces improve corrosion resistance of copper in seawater, Electrochim. Acta, *2007*, 52(11), 3709–3882

[172] Zhang F., Zhao L., Chen H., Xu S., Evans D.G., Duan X.: Corrosion resistance of superhydrophobic layered double hydroxide films on aluminium, Angew. Chem. Int. Ed., *2008*, 47(13), 2466–2469

[173] Zhang F., Chen S., Dong L., Lei Y., Liu T., Y.Yin: Preparation of superhydrophobic films on titanium as effective corrosion barriers, Appl. Surf. Sci., *2011*, 257(7), 2587–2591

[174] Rao A.V., Latthe S., Mahadik S., Kappenstein C.: Mechanically stable and corrosion resistant superhydrophobic sol–gel coatings on copper substrate, Appl. Surf. Sci., *2011*, 257(13), 5772–5776

[175] Weng C.J., Chang C.H., Peng C.W., Chen S.W., Yeh J.M., Hsu C.L., Wei Y.: Advanced anticorrosive coatings prepared from the mimicked Xanthosomasagittifolium leaf-like electroactive epoxy with synergistic effects of superhydrophobicity and redox catalytic capability, Chem. Mater., *2011*, 23(8), 2075–2083

[176] Yang T.I., Peng C.W., Lin Y.L., Weng C.J., Edgington G., Mylonakis A., Huang T.-C., Hsu C.-H., Yeh J.-M., Wei Y.: Synergistic effect of electroactivity and hydrophobicity on the anticorrosion property of room-temperature-cured epoxy coatings with multi-scale structures mimicking the surface of Xanthosomasagittifolium leaf, J. Mater. Chem., *2012*, 22(31), 15845–15852

[177] Hil R., Ma M.: Superhydrophobic surfaces, Curr. Opin. Colloid Interface Sci., *2006*, 11(4), 193–202

[178] Chang K.-C., Hsu M.-H., Lu H.-I., Lai M.-C., P Liu.-J., Hsu C.-H., Ji W.-F., Chuang T.-L., Wei Y., Yeh J.-M., Liu W.-R.: Room-temperature cured hydrophobic epoxy/graphene composites as a corrosion inhibitor for cold-rolled steel, Carbon, *2014*, 66(1), 144–153

[179] Turhan M.C., Li Q., Jha H., Singer R.F., Virtanen S.: Corrosion behaviour of multiwall carbon nanotube/magnesium composites in 3.5% NaCl, Electrochim. Acta, *2011*, 56(20), 7141–7148

[180] Li Q., Turhan M.C., Rottmair C.A., Singer R.F., Virtanen S.: Influence of MWCNT dispersion on the corrosion behaviour of their Mg composites, Mater. Corr., *2012*, 63(5), 384–387

[181] Fukuda H., Szpunar J.A., Kondoh K., Chromik R.: The influence of carbon nanotubes on the corrosion behaviour of AZ31B magnesium alloy, Corr. Sci., *2010*, 52(12), 3917–3923

[182] Aung N.N., Zhou W., Goh C.S., Mui S., Nai L., Wei J.: Effect of carbon nanotubes on the corrosion of Mg–CNT composites, Corr. Sci., *2010*, 52(5), 1551–1553

[183] Zhang M., Yan Y., Gong K., Mao L., Guo Z., Chen Y.: Electrostatic layer by layer assembled carbon nanotube mutilayer film and its catalytic activity for the oxygen reduction reaction, Langmuir, *2004*, 20(20), 8781–8785

[184] Jürmann G., Tammeveski K.: Electroreduction of oxygen on multi-walled carbon nanotube modified highly oriented pyrolytic graphite electrodes in an alkaline solution, J. Electroanal. Chem., *2006*, 597(2), 119–126

[185] Wang S., Iyyamperumal E., Roy A., Xue Y., Yu D., Dai L.: Vertically aligned BCN nanotubes as efficient metal-free electrocatalysts for the oxygen reduction reaction: A synergetic effect by co-doping with boron and nitrogen, Angew. Chem., *2011*, 123(49), 11960–11964

[186] Kumar C.M.P., Venkatesha T.V., Shabadi R.: Preparation and corrosion behavior of Ni and Ni–graphene composite coatings, Mater. Res. Bull., *2013*, 48(4), 1477–1483

[187] DENNIS R.V., VIYANNALAGE L.T., GAIKWAD A.V., ROUT T.K., BANERJEE S.: Graphene nanocomposite coatings for protecting low-alloy steels from corrosion, Am. Ceram. Soc. Bull., *2013*, 92(5), 18–24

[188] ROUT T.K., GAIKWAD A.V., DINGEMANS T.A.: A method of preparing a polyetherimide coating on a metallic substrate, World Intellectual Property Organization, Pat. No. WO2011035920, A1, *2011*

[189] ROUT G.K., GAIKWAD A.V., LEE V., BANERJEE S.: Hybrid nanocomposite coatings for corrosion protection of low carbon steel: A substrate-integrated and scalable active–passive approach, J. Mater. Res., *2011*, 26(6), 837–844

[190] ROUT T.K., JHA G., SINGH A.K., BANDYOPADHYAY N., MOHANTY O.N.: Development of conducting polyaniline coating: A novel approach to superior corrosion resistance, Surf. Coat. Technol., *2003*, 67(1), 16–24

[191] KUMAR S., SUN L.L., CACERES S., LI B., WOOD W., PERUGINI A., MAGUIRE R.G., ZHONG W.H.: Dynamic synergy of graphitic nanoplatelets and multi-walled carbon nanotubes in polyetherimide nanocomposites, Nanotechnology, *2010*, 21(10), 105702–105710

[192] WILSON D., STENZENBERGER H.D., HERGENROTHER P.M.: Polyimides, Chapman and Hall, London, U.K., *1990*

[193] KUMAR S., LI B., CACERES S., MAGUIRE R.G., ZHONG W.-H.: Dramatic property enhancement in polyetherimide using low-cost commercially functionalized multi-walled carbon nanotubes *via* a facile solution processing method, Nanotech., *2009*, 20(46), 465708–465716

[194] VON BAECKMANN W., SCHWENK W., PRINZ W. (Eds.): Handbook of corrosion protection, Theory and practice of electrochemical protection processes, Third Edition, Golf Publishing Company, Houston TX, USA, *1997*

INVESTIGATIONS OF BIO-GASOIL PRODUCTION

Péter Solymosi, ✉ Zoltán Varga, and Jenő Hancsók

MOL Department of Hydrocarbon- and Coal Processing, University of Pannonia, Egyetem u. 10.,
Veszprém, 8200, HUNGARY
✉Email: solymosip@almos.uni-pannon.hu

Liquid engine fuels are the main source of power for transportation in the passenger sector. It is the projection of the European Union (EU) to reach 10% utilisation of renewable fuels by 2020. To achieve this goal the EU created the 2003/30/EC and furthermore the 2009/28/EC Directives. For example, the feedstocks of these renewable engine fuels can be non-edible oil plant hybrids, such as rapeseed oils with high euric acid content obtained from special hybrids of rape (e.g. *Brassica napus*) waste lards (used cooking oil and slaughterhouse lards). If the preconditions of utilisation are given with respect to the sustainability and technical compatibility of motor engines and vehicle construction, these bio components can be blended with motor fuels in large quantities. Considering the properties of currently used first generation biofuels, the maximum amount of bio-component in engine fuels is approximately 7 (v/v)% fatty acid-methylester in diesel fuels. A reliable production technology of second generation biofuels, which can be blended into diesel fuels is the heterogenic catalytic hydrogenation of triglycerides and waste lards. Furthermore, isomerisation can improve the quality of a bio-paraffin mixture. In this context, we studied the isomerisation of bio-paraffin mixtures, which were obtained from the hydrodeoxygenation of vegetable oil. The characteristics of these products were favourable, such as their cetane number being higher than 75, for example. The actual EN590:2013 standard does not limit the blending ratio of the paraffinic bio-component in diesel fuels. Consequently, these products obtained by the catalytic hydrogenation of vegetable oils can be blended into gasoil by up to 10 % or even more to meet the above EU requirements with respect to the utilisation of renewable fuels.

Keywords: bio gasoil, hydrodeoxygenation, catalytic conversion, biofuels, blending diesel fuels

Introduction

Interest in alternative fuels is on the rise due to the unequal presence of the fossil energy carriers, the periodic rise in the price of fossil fuels, the need for decreasing dependence on crude oil, and the regulations of the European Union. They can play a significant role in achieving the EU plan to reach a 10% energy ratio of total fuel consumption using alternative fuels by 2020. Thus, the application of the biofuels can be increased to a large degree in the long- and medium-terms. For example, in some countries the domestic demand on biofuels could increase to 20% by 2030, along with the decrease in the demand for engine fuels that could be up to 70%. The world's energy production from biomass could reach 5% by 2050 [4, 5]. Accordingly, to ensure the availability of this feedstock the production costs could decrease. To achieve these goals, the EU created several directives (1998/70/EC, 2001/77/EC, 2003/17/EC, 2003/30/EC, 2003/87/EC, 2009/28/CE, and 2009/30/CE). Natural triglycerides like vegetable oils (edible or non-edible/waste) can be feedstock for biofuels as alternative energy sources [6, 7], such as special breeding non-edible oil plants [8, 9], animal fats or waste cooking oil [10, 11]. During the conversion of natural triglyceride molecules to bio-gasoil the following reactions take place [1, 2, 3]:

- full saturation of double bonds (hydrogenation),
- heteroatom removal
  - oxygen removal
    - hydrodeoxygenation (HDO reaction, and reduction)
    - decarboxylation,
    - decarbonylation
  - removing of other heteroatoms (sulphur, nitrogen, phosphorous, and metals),
- isomerisation of *n*-paraffins that are formed during the removal of oxygen
- different side reactions
  - hydrocracking of the fatty acid chain of triglyceride molecules,
  - water-gas shift reaction
  - methanisation,
  - cyclisation, aromatisation, etc.

During the HDO reduction reaction normal paraffins are formed with carbon numbers that are equal to the fatty acids in triglycerides. In the case of decarboxylation and decarbonylation reactions (HDC) normal alkanes are produced, where the carbon number is one less than that of fatty acids of the original vegetable (*Fig.1*).

Bio-gasoil is a mixture of gasoil with the boiling range of *iso-* and *normal*-paraffins. It can be obtained by the hydrogenation of vegetable oils and natural triglycerides
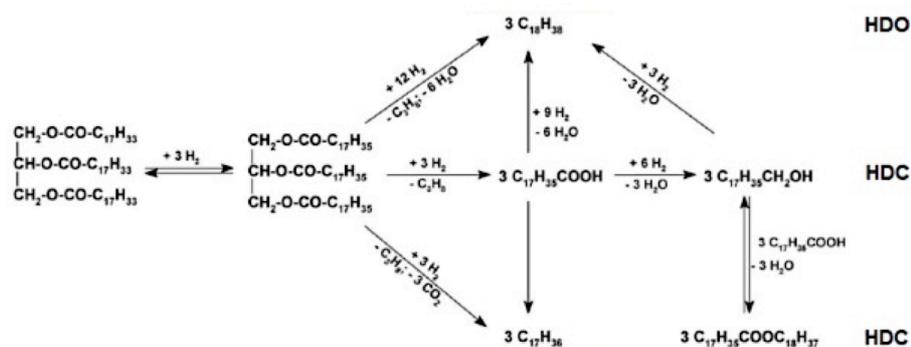
Figure 1: Pathways for the removal of oxygen from vegetable oils
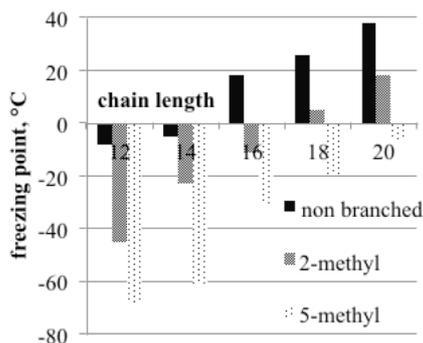
Figure 2: The freezing point of *iso*-paraffins as a function of the branch position

from other sources. These constitute the second generation biofuel components of diesel engines. They have good quality characteristics, such as high cetane number, good flow properties, unlimited mixability with engine fuels, and the a production line compatible with existing refinery structures [18, 19]. The actual EN 590:2013 standard does not limit the blending ratio of second generation bio-components, while the blending of biodiesel is limited to 7 v/v%. All the above mentioned aspects of alternative fuels can rationalise the investigation of the hydrogenation of non-traditional feedstock sources. These are the vegetable oils that can be obtained from non-edible hybrid oil-plants, rapeseed oil from *Brassica napus* with high euric acid content to produce diesel fuel blending components with good flow properties in colder conditions (below +5 °C). The freezing point of *iso*-paraffins from bio-sources is lower than for equal chain length *normal*-paraffins (*Fig.1*) [12-15, 17]. Thus, products with high *iso*-paraffin contents have more favourable cold flow properties (CFPP) with cloud points at lower temperatures (*Fig.2*). The aim of our work was the production of diesel gasoil blending components *via* the isomerisation of paraffin mixtures obtained from the hydrodeoxygenation of rapeseed oil with high euric acid content.

Experimental

In this work, a diesel gasoil bio-blending component production technique was investigated that meets the requirements of the EN:590 Standard with the possibility of blending it with engine fuels in unlimited quantities. Thus, the hydrodeoxygenation of natural

triglycerides and further the isomerisation of the obtained bio-paraffin mixture were investigated over the Pt-SAPO-11 catalyst [16] developed in-house. The effect of the operation parameters, such as temperature, pressure, and liquid hourly space velocity (LHSV) was studied on the yield, composition, and utilisation properties of the products.

Experimental Apparatus and Product Separation

The experimental tests were carried out in one of the measured sections of a high-pressure reactor system containing two tubular reactors with a isothermal catalyst volume of $100 \, \text{cm}^3$. The reactor system contained all the equipment and devices applied in the reactor system of a hydrotreating plant. The apparatus is suitable for maintaining if not succeeding the industrial precision of main process parameters.

Analytical Methods

The main properties of the feedstock materials and products were determined by standard methods. The hydrocarbon composition of the bio-paraffin mixture was determined by high temperature gas chromatography (Shimadzu 2010 GC [column: Phonomenex Zebron MXT]).

Process Parameters

The ranges of the applied process parameters in the isomerisation test on the basis of our earlier experimental results [13, 14, 17, 20-23] were as follows: temperature 300–360 °C, total pressure 20–80 bar, liquid hourly space velocity (LHSV) $1.0 \, \text{h}^{-1}$, and H_2/feed volume ratio of $400 \, \text{Nm}^3 \, \text{m}^{-3}$.

Feedstock materials

The feedstock of the catalytic tests was a bio-paraffin mixture, which was obtained from the hydrodeoxygenation of rapeseed with euric acid produced in Hungary. It was properly filtered as a pre-treatment. The main properties of the feedstock material are shown in *Table 1*. The catalyst was Pt-SAPO-11 (0.5 % Pt), the main properties of this can be found in *Table 2*.

Table 1: Selected properties of the feedstock materials

Properties	rapeseed oil	Bio-paraffin mixture		
kinematic viscosity at 40 °C, mm² s⁻¹	46.56	3.493		
density at 15 °C, g cm⁻³	0.9804	0.7923		
cloud point, °C	16	32		
cetane number	42	104		
compositions, %	Fatty acid		Paraffin	
	C16:0	2.3	C₁₄₋	0.2
	C16:1	0.1	C₁₄	0.1
	C18:0	1.2	C₁₅	0
	C18:1	28.8	C₁₆	2.3
	C18:2	12.4	C₁₇	29.5
	C18:3	8.3	C₁₈	28.8
	C20:0	0	C₁₉	6.1
	C20:1	4.8	C₂₀	5.6
	C22:0	0.1	C₂₁	14.8
	C22:1	41.8	C₂₂	12.5
	other	0.2	C₂₂₊	0.1

Table 2: Selected properties of the isomerisation catalyst used

Properties	Pt/SAPO-11
Pt content, w%	0.5
Pt dispersity, %	69
BET surface area, m² g⁻¹	105
average pore size, nm	0.61
micropore volume, cm³ g⁻¹	0.06
macropore volume, cm³ g⁻¹	0.20
total pore volume, cm³ g⁻¹	0.26
acidity, mmol NH_3 g⁻¹	0.13
acidity (rel.), mmol NH_3 m⁻² cat.	0.0012

Results and Analysis

The first step was to produce a bio-paraffin mixture with a boiling range of gasoil from rapeseed with a high euric acid content. The properties of the bio-paraffin are summarised in *Table 1*. The commercially available $NiMo/Al_2O_3$ catalyst was utilised for the production of the bio-paraffin mixture. During the catalytic test the employed operation parameters were as follows: 320–380 °C, 20–80 bar, LHSV = 1.0 h⁻¹, and H_2/CH ratio of 600 Nm³ m⁻³ [8]. It was found that the favourable operation parameters are 340 °C, 40 bar, LHSV=1.0 h⁻¹, and H_2/CH ratio of 600 Nm³ m⁻³. The tested catalyst is suitable for the production of bio-paraffin mixtures with

high yields from natural triglycerides. Due to the moderate acidity of this catalyst, the formation of *iso*-paraffins was lower (5 wt%, *Fig.3*). Accordingly, the CFPP of the products was found to be high (27 °C). The product fraction produced in this way, in practice, cannot be blended into diesel fuels in low temperate zone countries. It is necessary then for the improvement of CFPP *via* the catalytic isomerisation of this mixture with high *normal*-paraffin content [10, 11]. A large amount of bio-paraffin mixture was produced in a thousand hour, long-term catalytic test. The target fraction of the isomerisation tests was the 180–360 °C boiling range, which is the boiling range of gasoil. The yield of the target products was higher than 94 % in all operation parameter combinations (*Fig.4*). The lighter fraction with a boiling range of up to 180 ºC contains mainly *iso*-paraffins, which can be outstanding gasoline blending components due to their high octane numbers (>85).

We found that by adjusting the operation parameters, such as increasing the temperature, and decreasing the LHSV, the yield of the target fraction was decreased due to the higher yield of the cracking reaction. The target fraction obtained between 70% and 80%
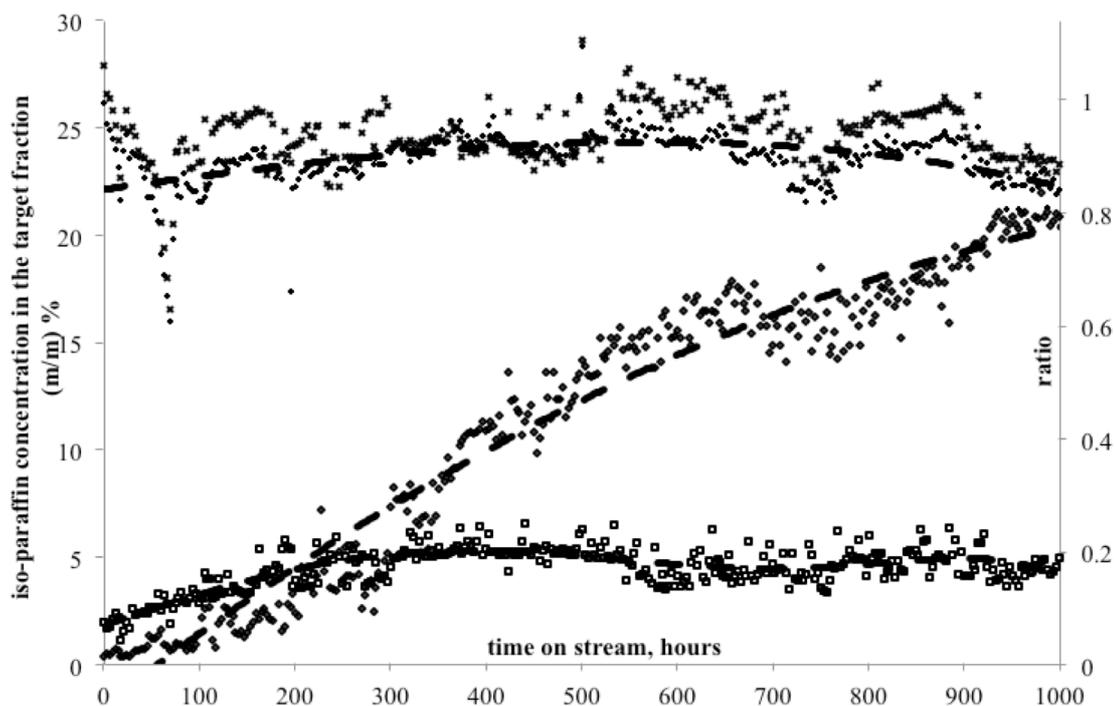
Figure 3: Hydrogenation of rapeseed oil with high euric acid content
(diamond: residual triglyceride, square: *iso*-paraffin content, cross C_{21}/C_{22} ratio)
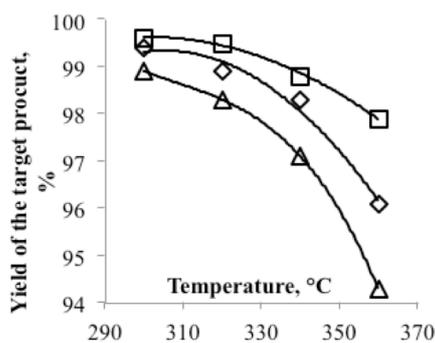
Figure 4: The yield of the target fraction as a function of operation parameters (pressure: 40 bar, liquid hourly space velocity square: 1.0 h^{-1}, diamond 2.0 h^{-1}, triangle 3.0 h^{-1})
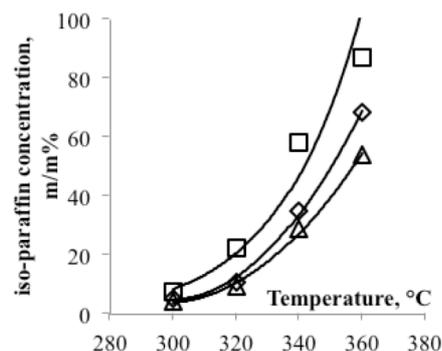
Figure 5: The iso-paraffin concentration of the target fraction as a function of operation parameters (pressure: 40 bar, liquid hourly space velocity: square 1.0 h^{-1}, diamond 2.0 h^{-1}, triangle 3.0 h^{-1})
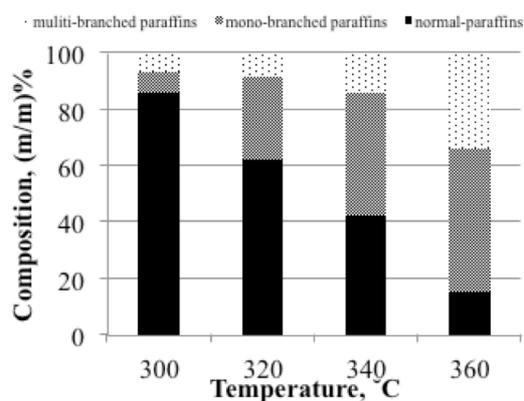
Figure 6: The composition of the products as a function of operation parameters (pressure: 40 bar, liquid hourly space velocity: 1.0 h^{-1})
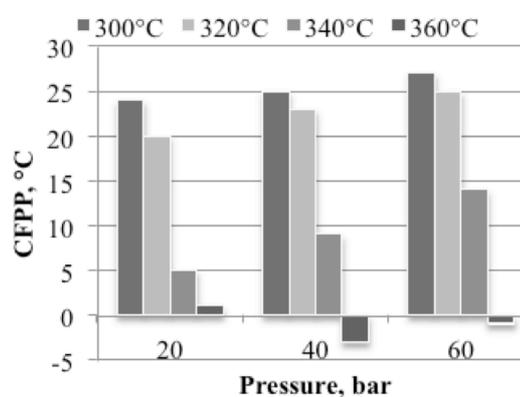
Figure 7: CFFP of the products as a function of operation parameters (liquid hourly space velocity: 1.0 h^{-1}, H$_2$/feed ratio: 400 Nm^3 m^{-3})

contained C_{17}–C_{22} hydrocarbons, as well as other (C_{13}–C_{16}) hydrocarbons from the boiling range of gasoil. The iso-paraffin content of the target fraction increased significantly with the operating temperature (Fig.5). The increase of the iso-paraffin concentration occurred at 360 °C then at higher temperatures it started to decrease, due to the thermodynamic hindrance of the exothermic reactions, and the higher rate of cracking reactions.

Up to ca. 320 °C, mainly mono-branched iso-paraffins were formed and were by in large mono-ethyl-paraffins (Fig.6). The freezing points of these products are much lower than normal-paraffins and the cetane number is high enough for a fuel additive. The greater formation of mono-methyl-paraffins over the SAPO-11 catalyst can be explained by the reduced formation of iso-paraffins due to steric hindrance. At 340 °C or higher, the formation of multi-branched isomers was significant (Fig.6). These compounds have better cold flow properties (below -20 °C), but their cetane numbers are high enough (30–45) as shown in Fig.2. The favourable operation parameters in terms of bio-gasoil yield and iso-paraffin concentration were as follows: T = 360 °C; p = 40 bar; and H$_2$/feedstock ratio = 400 Nm^3 m^{-3}. The CFPP values of the products as a function of temperature and operation pressures are shown in Fig.7. These components have low enough CFPP values to blend into diesel gasoil in moderate amounts. On the basis of the experimental results, it was concluded that the production of bio-gasoil meets the standard's requirements with a CFPP value of max. +5 °C and 70% iso-paraffin content (Fig.8) in the case when the raw material contains 8% C_{17}–C_{22} iso-paraffins.

Conclusions

Based on our experimental results, it was concluded that the NiMo/Al$_2$O$_3$ catalyst is suitable for the long-term production of bio-paraffin mixtures from natural triglycerides via catalytic hydrodeoxygenation. Furthermore, the investigated Pt-SAPO-11 catalyst is suitable for improving the quality of a bio-paraffin mixture that was obtained from the hydrodeoxygenation of rapeseed oil with high euric acid content. During the isomerisation with optimised operation parameters, the yield of the target fraction was higher than 94%. At 340 °C or higher the iso-paraffin content is close to 70%. Consequently, the cold flow property of the cloud point is lower than +5 °C. Therefore, this approach can produce gasoil bio-blending components with good utilisation properties, such as high cetane number, and low temperature values for cold flow properties. Overall, the products described here are suitable for blending components of diesel fuels with concentrations of 10% or higher.
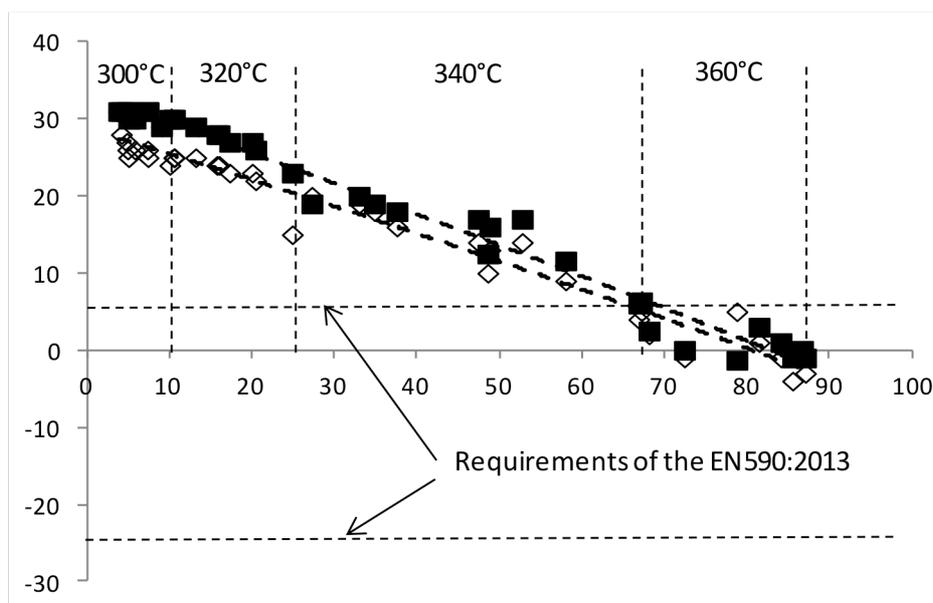
Figure 8: Cold flow properties as a function of *iso*-paraffin concentration
(solid squares: cold filtering clugging point, hollow diamonds: cloud point)

REFERENCES

[1] HUBER G.W., O'CONNOR P., CORMA A.: Processing biomass in conventional oil refineries: Production of high quality diesel by hydrotreating vegetable oils in heavy vacuum oil mixtures, Appl. Catal. A:General, *2007*, 329,120-129

[2] DA ROCHA FILHO G.N., BRODZKI D., DJEGA-MARIADASSOU G.: Formation of alkanes, alkylcycloalkanes and alkylbenzenes during the catalytic hydrocracking of vegetable oils, Fuel, *1993*, 72(4), 543-549

[3] SIMACEK P., KUBICKA D., SEBOR G., POSPISIL M.: Fuel properties of hydroprocessed rapeseed oil, Fuel, *2010*, 89(3), 611-615

[4] SCHOLLNBERGER W.E.: Between 2020 and 2030 to be re-joined the course for the global energy mix, Erdöl Erdgas Kohle, *2013*, 129(12), 2-7 (in German)

[5] Shell Deutschland Oil Gmbh: Shell study sees great potential for biofuels, Erdöl Erdgas Kohle, *2012*, 128(11), 407-411 (in German)

[6] KRÁR M., KOVÁCS S., KALLÓ D., HANCSÓK J: Fuel purpose hydrotreating of sunflower oil on CoMo/Al$_2$O$_3$ catalyst, Bioresources Technology, *2010*, 101(23), 9287-9293

[7] KRÁR M., KASZA T., KOVÁCS S., KALLÓ D., HANCSÓK J.: Bio-gasoils with improved low temperature properties, Fuel Proc. Techn., *2011*, 92(5), 886-892

[8] SOLYMOSI P., BALADINCZ P., HANCSÓK J.: Production of motor fuels by the hydrogenation of rapeseed oil with high euric acid content, Proc. 19th EU Biomass Conference and Exhibition, Berlin, Germany, *2011*, 2077-2081

[9] SOLYMOSI P., KASZA T., HANCSÓK J.: Investigation of conventional and high oleic acid content rapeseed or sunflower oils, Hung. J. Ind. Chem., *2011*, 39(1), 85-90

[10] BALADINCZ P., TÓTH CS. HANCSÓK, J.: Expanding feedstock supplies of the second generation biofuels of diesel engines, Hung. J. Ind. Chem., *2010*, 38(1), 1-7

[11] SOLYMOSI P., ELLER Z., HANCSÓK J.: Motor fuel purpose hydrogenation of used cooking oils, Chem. Engng. Trans., *2013*, 35(2), 1351-1356

[12] HANCSÓK J., KASZA T., KOVÁCS S., SOLYMOSI P., HOLLÓ A.: Production of bioparaffins from natural triglycerides, Chem. Engng. Trans., *2011*, 25, 821-826

[13] KASZA T., HANCSÓK J.: Isomerisation of bio-paraffins over Pt-based SAPO-11 and AlMCM-41 catalysts, Műszaki Kémiai Napok, *2011*, 29, 220-225 (in Hungarian).

[14] KASZA T., HOLLÓ A., THERNESZ A., HANCSÓK J.: Production of bio-gasoil from bioparaffins over Pt/SAPO-11, Chem. Engng. Trans., *2010*, 21, 1225-1230

[15] HANCSÓK J., KASZA T., KOVÁCS S., SOLYMOSI P., HOLLÓ A.: Production of bio-gasoil from bioparaffins over Pt/SAPO-11, J. Cleaner Prod., *2012*, 34, 76-81

114

[16] GERGELY J., FORSTNER J., BALAI M., SZIRMAI L., PETRO J., KOVÁCS I., RESOFSZKI G., SMID L., AUER J., BALADINCZ J., SZALMÁSNÉ PÉCSVÁRI G., TÓTH L., KUBOVICSNÉ STOTZ K., KUN J., KÁNTOR L., CZÁGLER I., LŐVEI J., TÓTH E., LEFKÁNICS GY., TÁTRAI E., PÁLNÉ BORBÉLY G., ROSENBERGNÉ MIHÁLYI M., BEYER H., HANCSÓK J.: A selective hydroisomerisation catalyst preparation and application for hydrocarbons HU 225 912, *2001* (in Hungarian)

[17] KASZA T., KALLÓ D., HANCSÓK J.: Quality improvement of bio-paraffin mixtures, Fuel, *2014*, 120, 1-7

[18] HANCSÓK J., KRÁR M., MAGYAR SZ., BODA L., HOLLÓ A., KALLÓ D.: Investigation of the production of high cetane number biogasoil from pre-hydrogenated vegetable oils over Pt/HZSM-22/Al$_2$O$_3$, Micropor. Mesopor. Mat., *2007*, 101(1-2), 148-152

[19] SRIVASTAVA S.P., HANCSOK J.: Fuels and Fuel Additives, Wiley and Sons, Hoboken, NJ, USA, 2014, pp. 121-177

[20] KASZA T., BALADINCZ P., HANCSÓK J.: Production of bio-isoparaffins by the hydroisomerisation of bioparaffins. Hung. J. Ind. Chem., *2009*, 37(2), 95–99

[21] KASZA T., HANCSÓK J.: Production of depressed freezing point bio-gasoil from slaughterhouse waste lard. Hung. J. Ind. Chem., *2010*, 38(1), 41–45.

[22] KASZA T., HANCSÓK J.: Isomerisation of paraffin mixtures produced from sunflower oil. Hung. J. Ind. Chem., *2011*, 39(3), 363–368

[23] KASZA T.; TÓTH CS.; HANCSÓK J.: Application of improved bio-paraffins in diesel fuels, Hung. J. Ind. Chem., *2012*, 40(1), 25–31

THE ENERGY BALANCE OF SEPARATION OPPORTUNITIES
IN MICROALGAE TECHNOLOGIES

Zoltán Hodai, ✉ Dóra Rippel-Pethő, Géza Horváth, László Hanák, and Róbert Bocsi

Department of Chemical Engineering Science, University of Pannonia,
Veszprém, Egyetem u. 10., 8200, HUNGARY
✉Email: hodaiz@almos.uni-pannon.hu

Algae technology is at the focus of international research and development, since it is a green technology that reduces emissions of harmful chemicals and can be considered as a renewable energy source. Carbon dioxide from stack gases and the nitrogen content of wastewater can be considered as food sources for plants and algae. The utilisation of carbon dioxide by algae technologies depends on the technical environment and logistics of teamwork. This technology is a new opportunity in Hungary for decreasing emissions. We grew algae populations to utilise the carbon dioxide from a refinery's stack gas in the continental climate of Hungary. Critical parameters of the technology are the concentration of the algae suspension and extract, because of high investment and operating costs as well as the long operation time, which determines the feasibility of the algae technology. Our specific aim was to separate the algae mass faster and more efficiently from the starting solution. The optimisation of separation operations and technologies took into consideration environmental and economic aspects.

Keywords: microalgae technology, separation, filtration, renewable energy source

Introduction

The utilisation of microalgae carbon dioxide fixation is an important area of international research and development. The absorption of certain technological exhaust gases is possible on the basis of the photosynthesis of microalgae. The absorption of carbon dioxide can reach a magnitude of hundred ton per hectare. C_{16}-C_{22} esters are formed in certain algae cells that can be used for the production of biodiesels. This method is thus capable of producing fuels from renewable sources [1-4].

Algae production is a promising solution amongst the alternative fuel production processes, because it requires of low specific area for growth and high reproduction rate [5-8]. Algae are considered to be one of the most efficient organisms on Earth due to their outstanding reproduction rate, and generally high lipid content. For example, they can double their biomass in 24 hours [9-12]. Their lipid content on average is 20%, but it can be up to 60-80% for certain species [7, 9, 13, 14].

Research into oil production from algae is primarily based on microalgae. These are photosynthesising organisms with a cell size of no greater than 0.5 mm. They can be utilised for carbon dioxide and nitrogen oxide fixation, because they convert these compounds in a photosynthetic energy conversion [15]. The end product of these processes contains a significant amount of solar energy stored as chemical energy.

Furthermore, considerable amounts of biodiesel can be obtained [16-19]. The composition of a microalgae cell depends on cultivation parameters. We tested the available and applicable species under local climatic conditions. Afterwards microalgae that passed the local environmental tests can be considered useable for production.

Research is being carried out into carbon dioxide fixation from technological flows at our institute [15]. The absorption of carbon dioxide and reduction of the release of other pollutants in wastewater using microalgae are being studied. Algae technology utilises waste gases and some environmentally harmful components from wastewaters as nutrients, and thus purifies the growth media. These specific pollutants provide excessive amount of nutrients for the algae, which results in the algae's exponential growth. In addition to the above-mentioned method of energy extraction, a number of research efforts [4, 5, 8, 10-12, 18, 19 25] are currently dealing with the alternative use of biomass produced in this way, and biomass residue that remains after processing. In addition to the above-mentioned advantages, the operating costs can be a limiting factor.

The most critical steps in the production of algae-based biofuels are the harvesting of algae (harvesting, dewatering, and drying), and lipid extraction, because of the high level of investment and operating costs. The main challenge of the technology is to reduce costs, which by in large originate from the separation steps that need to be minimised.

Figure 1: Closed grower systems at the University of Pannonia

Figure 3: Open grower systems at a refinery

Results and Discussions

Our research focuses on carbon dioxide fixation from technological flows. We designed and built various grower systems at the Department of Chemical Engineering of the University of Pannonia (*Fig.1*), as well as at a refinery (*Figs.2* and *3*). By utilising these technological solutions we conducted research into the production of biomass and algae-based products as possible renewable fuels.

The propagation and environmental tolerance parameters were examined along with the possibilities of developing the technology. During the experiments, the whole technological chain was examined providing the possibility to optimise the entire chain of operational steps. The utilisation of algae cultures in experimental photobioreactors is examined, together with the optimisation of the operational conditions both for both artificial and natural light with different substrate solutions. The various parameters for algae processing are also determined.

The foci of our work were the processing and separation operations. Critical points of the technology are the processing steps, such as concentrating the algae suspension and extracting valuable components (lipids). The extraction technologies are of importance because of the high costs of investment and long operation periods. From the literature [5, 10, 12, 15, 17, 22, 23], biodiesel is not yet comparable to fuel produced from petroleum, however the cost of algae technology is dropping. Furthermore, algae technology could become viable if we consider the cost of wastewater purification and flue gas adsorption, and the price of products obtained from microalgae. We need to consider

Figure 2: Closed grower systems at a refinery

separating possibilities and beyond by analysing gains and losses simultaneously.

Our specific aim was to devise densification and separation processes, which have low energy requirements and advantageous operation times. Furthermore, we defined useful components from algae and their optimised extraction, based on the optimisation of extraction techniques and other economical and environmental aspects.

Separation Opportunities

Harvesting can be carried out by microfiltration, ultra filtration, centrifugation, flocculation, sonochemical techniques, or some new techniques that are under development [21, 22]. In addition to chemical flocculation, clarification, and membrane separation procedures, special attention was paid to auto flocculation phenomena.

Forms of technology pay more and more attention to convert disposed waste into useful materials. The photosynthesising microorganisms, such as microalgae, utilise solar energy, rapidly reproduce, and do not require any soil to grow in. The biomass product contains solar energy stored in chemical bonds. To process the final product, the suspension, concentration, and extraction of the biomass are the most problematic parts of the methods of algae-based energy production. The main techniques for separation are mechanical operations (filtration, centrifugation, and settling), mechanical operations with admixture (flocculation and defecation), membrane operations (microfiltration and ultrafiltration), and other notable operations (sonochemical techniques, electroflocculation, and flotation). According to examples from the literature [9, 11, 15, 18], there are still only limited generally applicable and proven methods that can be used for biomass production with specific energetic goals.

Energetic Considerations

Increasingly diluted suspensions were examined during separation operations (*Fig.4*). Data from the literature [23-37] need to be brought to a common denominator for carrying out comparisons. For example, our results show that if we want to gradually treat more dilute suspensions using flotation, the energetic considerations can change by orders of magnitude.
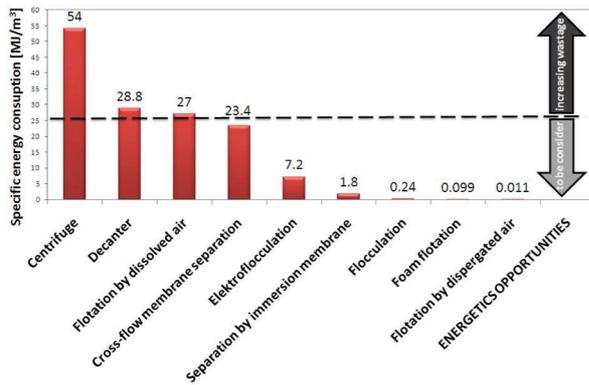
Figure 4: The energetic consideration of selected separation processes (the dashed line shows the energy content of algae suspensions on average of 25 MJ m⁻³)

According to *Fig.4*, cross-flow membrane separation, electroflocculation, centrifugation known as "spiral plate", immersion membrane separation, the flocculation, foam flotation and operation of the flotation by dispersed air are positioned favourably with respect to energetic rating (smaller demand of energy as the energy amount from the separated biomass).

The energy balances of chemical flocculation and electroflocculation do not include the costs of the procurement of chemicals and post-treatment, which would complicate the determination of its energy status. The amounts of chemicals used according to the quality of the suspension were between such wide intervals that they could not be considered using simple factors. By energetic rating, we should note the concentration of the kind of algae suspension in the chosen operation. The increase in volume and decrease in concentration of the suspension may require a review of the cost, materials and energy needs of the separation, and thus change its energetic status. For example, we can conduct a thought-experiment for flotation to see whether the treatment of a dilute suspension needs a device with a different energetic rating. Because of the fixed size of the flotation's device, we cannot decrease the amount of gas flow and the dilute suspension could start to foam, which may make it more difficult to handle than the stability and volume of the foam from a more concentrated suspension. Due to foaming, it will be more difficult for the algae layer to become thicker in the foam unless we use more chemicals and surface-active agents, which result in more parameters to consider. Apart from the above, an important question is how we consider the time domain of the technology (periodic, half-continuous, or continuous). The kind of growing system determines how to connect the separation operation. Half-continuous or continuous (or almost that) operation is beneficial for industrial biomass production to maximise the capacity of biomass. By running the reactor at the maximum rate of reproduction, we can maximise its capacity. According to the latter, the immersion membrane separation operation would be preferable.

Conclusions

According to our investigations, the usable operations for extraction and concentration in algae cultivation with regards to energetic causes vary greatly with respect to energy demands. For dilute suspensions, flotation and foam flotation are the most useful separation processes. We can use flocculation as well, but we have to consider the costs of and environmental damage caused by waste chemicals.

Acknowledgements

REFERENCES

[1] STRAKA F., DOUCHA J., LÍVANSKY K.: Utilisation of flue gas for cultivation of microalgae (Chlorella sp.) in an outdoor open thin-layer photobioreactor, J. Appl. Phycology, *2005*, 17(5), 403-412

[2] OLAIZOLA M.: Microalgal removal of CO_2 from flue gases: Changes in medium pH and flue gas composition do not appear to affect the photochemical yield of microalgal cultures, Biotechn. Bioproc. Engng., *2003*, 8(6), 360-367

[3] LAMENTI G., PROSPERI G., RITORTO L., SCOLLA G., CAPUANO F., PEDRONI P.M., VALDISERRI M.: Enitecnologie R&D project on microalgae biofixation of CO_2: outdoor comparative tests of biomass productivity using flue gas CO_2 from a NGCC power plant, Proc. 7th Int. Conf. Greenhouse Gas Control Technologies, 2005, 2(1), 1037-1042

[4] CARLSSON A.S., BILEN J.B., MÖLLER R., CLAYTON D.: Mircro- and macroalgae: utility for industrial applications, Ed.: BOWLES D., Outputs from the EPOBIO project. CPL Press, Berks, UK, *2008*

[5] BRIGGS M., VASUDEVAN P.T.: Biodiesel production-current state of the art and challenges, J. Ind. Microbiol. Biotechnol., *2008*, 35, 421-430

[6] HWANG E.J., SHIN H.S., CHAE S.R.: Single cell protein production of *Euglena gracilis* and carbon dioxide fixation in an innovative photobioreactor, Bioresource Technol., *2006*, 97(2), 322-329

[7] BECKER E.W., BADDILEY J.: Microalgae: Biotechnology and Microbiology, Cambridge Univ. Press, New York, USA, *1994*, p. 178

[8] POSEWITZ M.C., JINKERSON R.E., SUBRAMANIAN V.: Improving biofuel production in phototrophic microorganisms with systems biology tools, Biofuels, *2011*, 2(2), 125-144

[9] KOJIMA E., ZHANG K.: Growth and hydrocarbon production of microalgae *Botryococcus braunii* in bubble column photobioreactors, J. Biosci. Bioengng., *1999*, 87(6), 811-815

[10] BRIGGS M.: Wide-scale Biodiesel Production from Algae, Ph.D. Dissertation, University of New Hampshire, *2004*

[11] HWANG E.J., SHIN H.S., CHAE S.R.: Single cell protein production of *Euglena gracilis* and carbon dioxide fixation in an innovative photobioreactor Bioresource Technol., *2006*, 97(2), 322-331

[12] CHISTI Y.: Biodiesel from microalgae, Biotechnol. Adv., *2007*, 25(3) 294–306

[13] DISMUKES G.C.: Algal Photosynthesis, Princeton University Press, Princeton, NJ, USA *2008*

[14] SHI D., SONG D., FU J.: Construction of a shuttle vector for heterologous gene expression in *Escherichia coli* and microalgae anabaena, Chin. J. Biotechnol., *2008*, 24(3), 341-348

[15] BOCSI R., HORVÁTH G., HANÁK L.,: Microalgae production in service of fuel production, Hung. J. Ind. Chem., *2010*, 38(1), 9-13

[16] OLAIZOLA M., MASUTANI S.M., NAKAMURA T.: Recovery and sequestration of CO_2 from stationary combustion systems by photosynthesis of microalgae, U.S. Department of Energy, Office of Fossil Energy National Energy Technology Laboratory, Pittsburgh, PA, USA, *2006*

[17] BENEMANN J., SHEEHAN J., ROESSLER P., DUNAHAY T.: Biodiesel from algae, a look back at the U.S. DOE's aquatic species program, NREL Report NREL/TP-580-24190, *1998*

[18] BURLEW J.: Algae culture: from laboratory to pilot plant, Carnegie Institute, Washington DC, USA *1953*

[19] JUNG I.H., CHOE S.H.: Growth inhibition of freshwater algae by ester compounds released from rotten plants, J. Ind. Engng. Chem., *2002*, 8(4), 297-304

[20] SHELEF G.A., SUKENIK A., GREEN M.: Microalgae harvesting and processing: a literature review, Techn. Rep., Solar Energy Research Institute, *1984*

[21] POELMAN E., PAUW N.D., JEURISSEN B.: Potential of electrolytic flocculation for the recovery of micro-algae, Res. Conserv. Recyc., *1997*, 19(1), 1-10

[22] LEITE G.B., ABDELAZIZ A.E.M., HALLENBECK P.C.: Algal biofuels: challenges and opportunities, Bioresource Technol., *2013*, 145, 134-141

[23] RAWAT I., RANJITH KUMAR R., MUTANDA T., BUX F.: Biodiesel from microalgae: A critical evaluation from laboratory to large-scale production, Appl. Energy, *2013*, 103, 444-467

[24] DASSEY A.J., THEEGALA C.S.: Harvesting economics and strategies using centrifugation for cost-effective separation of microalgae cells for biodiesel applications, Bioresource Technol., *2013*, 128(2), 241-245

[25] UDOM I., ZARIBAF B.H., HALFHIDE T., GILLIE B., DALRYMPLE O., ZHANG Q., ERGAS S.J.: Harvesting microalgae grown on waste water, Bioresource Technol., *2013*, 150, 513-522

[26] UDUMAN N., BOURNIQUEL V., DANQUAH M.K., HOADLEY A.F.A.: A parametric study of electrocoagulation as a recovery process of marine microalgae for biodiesel production, Chem. Eng. J., *2011*, 174(1), 249-257

[27] BEACH E.S., ECKELMAN M.J., CUI Z., BRENTNER L., ZIMMERMAN J.B.: Preferential technological and life cycle environmental performance of chitosan flocculation for harvesting of the green algae *Neochloris oleoabundans*, Bioresour. Technol., *2012*, 121, 445–449

[28] BANERJEE C., GHOSH S., SEN G., MISHRA S., SHUKLA P., BANDOPADHYAY R.: Study of algal biomass harvesting using cationic guar gum from the natural plant source as flocculent, Carbohydrate Polymers, *2013*, 92(1), 675-681

[29] SCHLESINGER A., EISENSTADT D., BAR-GIL A., CARMELY H., EINBINDER S., GRESSEL J.: Inexpensive non-toxic flocculation of microalgae contradicts theories; overcoming a major hurdle of bulk algal production. Biotechnol. Adv., *2012*, 30(5), 1023–1030

[30] JUNGMIN K., BYUNG-GON R., KYOCHAN K., BO-KYONG K., JONG-IN H., JI-WON Y.: Continuous microalgae recovery using electrolysis: Effect of different electrode pairs and timing of polarity exchange, Bioresource Technol., *2012*, 123(2), 164-170

[31] MASCIA M., VACCA A., PALMAS S.: Electrochemical treatment as a pre-oxidative step for algae removal using Chlorella vulgaris as a model organism and BDD anodes, Chem. Enging. J., *2013*, 219(5), 512-519

[32] LEE A.K., LEWIS D.M., ASHMAN P.J.: Harvesting of marine microalgae by electroflocculation: The energetics, plant design, and economics, Appl. Energy, 2013, 108(3), 45-53

[33] AMER L., ADHIKARI B., PELLEGRINO J.: Techno-economic analysis of five microalgae-to-biofuels processes of varying complexity, Bioresource Technol., *2011*, 102(20), 9350–9359

[34] EDZWALD J.K.: Algae, bubbles, coagulants, and dissolved air flotation, Water Sci. Technol., *1993*, 27(10), 67–81

[35] NURDOGAN Y., OSWALD W.J.: Tube settling of high-rate pond algae, Water Sci. Technol., *1996*, 33(7), 229–241

[36] COLLET P., HÉLIAS A., LARDON L., RAS M., GOY R.A., STEYER J.P.: Life-cycle assessment of a microalgae culture coupled to biogas production, Bioresoure Technol., *2011*, 102(1), 207–214

[37] COWARDA T., LEEA J.G., CALDWELLB G.S.: Development of a foam flotation system for harvesting microalgae biomass, Algal Research *2013*, 2(2), 135-144

SOLID-LIQUID EXTRACTION OF CHLOROPHYLL FROM MICROALGAE FROM PHOTOAUTOTROPH OPEN-AIR CULTIVATION

ÉVA MOLNÁR, ✉ DÓRA RIPPEL-PETHŐ, AND RÓBERT BOCSI

Department of Chemical Engineering, Institute of Chemical and Process Engineering, University of Pannonia, Egyetem u. 10., Veszprém, 8200, HUNGARY
✉Email: molnare@almos.uni-pannon.hu

Among industrial pollutants, strict quotas limit the emission of carbon dioxide in the European Union. The capturing and deposition of carbon dioxide requires significant expenditures. One of the newer solutions for the reduction of the carbon dioxide emissions is provided by algae technology. In this technology, the absorption of carbon dioxide is achieved by photosynthesis. Besides the reduction of pollutants, algae technology has another advantage by supplying valuable products, such as natural pigments, proteins, vitamins and oils from algae biomass. Since it has a high reproduction rate, algae cultivation can be a feasible substitute for plants traditionally used in the production of chlorophyll. The viability of the technology is dependent on whether or not the processing can be done economically. An investigation was carried out in order to compare and contrast two extraction methods (Soxhlet extraction and leaching) and four solvents (acetone, diethyl ether, ethanol and methanol) to determine the most effective method for the extraction of chlorophylls from dried microalgae (*Chlorella vulgaris sp.*). It was concluded that methanol is the most effective solvent for the extraction of both chlorophylls *a* and *b* using both Soxhlet extraction and leaching.

Keywords: microalgae, extraction solvent, extraction, chlorophyll, spectrophotometry

Introduction

Energy demand has been rapidly increasing throughout the world, which is mainly met by the combustion of fossil fuels that results in an increase in the concentration of greenhouse gases in the atmosphere. However as a possible remediation strategy, carbon dioxide levels can be reduced with the use of algae technology [1]. Algae technology is of significant interest in research and development since it is a 'green' technology that is capable of both decreasing the emission of pollutants and serving as a renewable energy source. The algae absorb carbon dioxide for photosynthesis while producing a number of valuable components [2]. Microalgae are a collection of various microscopic species capable of photosynthesis and are typically found in water with an exceptionally high reproduction rate. For propagation, they mainly need light, water and CO_2. Some species can even survive in waste water. Algae are good alternatives for plants that are traditionally used for the production of chlorophyll. Appropriate conditions are required for the successful cultivation of algae (solar energy, temperature, pH, and mixing) [3, 4]. However, the critical point of algae technology is neither cultivation nor processing. The biggest complication concerns the concentration of the microalgae suspension and the subsequent extraction of the valuable components due to the high investment costs and long operational times [5]. There is, however, an increasing demand for microalgae. The reason for

this is that their oil content can be as much as 50% of their body mass, making them important in the production of biodiesel. The natural pigments, proteins and vitamins extracted from microalgae are primarily used by the pharmaceutical, cosmetic and food industries [3, 6].

Chlorophyll has been used for centuries as a traditional remedy for unpleasant body odours, for the neutralisation of the odours of stool and urine, and detoxification or sterilisation of wounds. Nowadays, its use is even more widespread. It is used as a food additive, food colouring, and nutritional supplement, especially because of its detoxifying and excellent basifying effects. It is also a strong antioxidant, which inhibits the harmful oxidative processes in the body and enhances the protection of cells and tissues [7, 8]. However, the processing of algae still needs to be improved with respect to determining the optimal methods, pieces of equipment, solvents, and parameters for the efficient and economic extraction of chlorophyll. Extraction is a process that can be carried out in a number of ways. There are also a number of solvents and preparation methods available. Also, variables such as pressure, temperature, the efficiency of contact and time all play an important role [9, 10]. Chlorophyll is sensitive to extreme light exposure, pH values and temperatures [7-9]. When choosing the solvent, one has to consider the sensitivity of the extracted component. A number of considerations are important, too: density, viscosity, heat of evaporation, price, effect on the environment and health. It is also essential that the
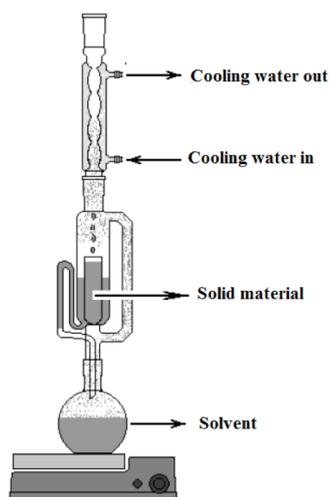
Figure 1: Soxhlet extractor [12]

solvent does not react with or cause damage to the extracted component and is not corrosive. Different methods can be used for different microalgae species [9-10].

Experimental

In advance of the extraction experiments the algae suspension (*Chlorella vulgaris sp.*) was concentrated, dried at 60 ℃ until it reached a constant mass, then ground in a ball mill. Chlorophylls *a* and *b* products were obtained by the methods detailed in the following sections from this ground algae powder, which contains approximately 4 wt% water.

The Soxhlet Extraction Method

One of the most well known pieces of equipment for laboratory scale solid-liquid extraction is the Soxhlet extractor. In a Soxhlet extractor multiple fractional distillations are carried out, the extraction is always done with the pure condensate [10].

The first step of the extraction was to fill a cellulose casing with 1-3 g of algae powder. The filled casing was placed in the middle part of the piece of equipment (*Fig.1*). The lower round-bottom flask was filled with 400 cm^3 of solvent. The Soxhlet extraction was carried out with methanol, ethanol, and acetone. Pumice and a magnetic stirring rod were placed into the lower flask to ensure proper boiling. The cooling water was set to a continuous flow and then the heating was turned on. After the extracting solvents had reached their boiling points and started to evaporate, they condensed in the reflux condenser and dripped back down onto the algae powder at which point the extraction of the chlorophylls started. The liquid level continuously rose in the middle section until it reached the overflow pipe letting the chlorophyll extract pour back into the lower flask. A sample of 4–6 cm^3 was taken at the end of every cycle. The extraction continued as long as the absorption spectra measured by the spectrophotometer did not show a significant change.

The Method of Leaching

Leaching was carried out with so-called 'cold solvents' at room temperature. Samples of 0.1 g; 0.5 g; and 0.75 g of algae powder were measured into a test tube, then 5–5 cm^3 of solvent was added. The list of solvents was the same as for the Soxhlet extraction method with the inclusion of diethyl ether. The weight of the solvent was also measured and recorded in order to have reference data in the later phases of the experiment. The samples were mixed with Vortex and centrifuged for 2 minutes at 4000 rpm. A sample was taken from the top layer after centrifuge and its absorbance measured with a spectrophotometer [11, 12].

Calculations

The chlorophyll contents of the given samples were calculated with empirical formulae according to *Eqs.(1)-(8)*.

$$c_{chlorophyll\text{-}a}(90\% \text{ methanol}) = 15.65 \cdot A_{666} - 7.34 \cdot A_{653} \quad (1)$$

$$c_{chlorophyll\text{-}b}(90\% \text{ methanol}) = 27.05 \cdot A_{653} - 11.21 \cdot A_{665} \quad (2)$$

$$c_{chlorophyll\text{-}a}(96\% \text{ ethanol}) = 13.95 \cdot A_{665} - 6.88 \cdot A_{649} \quad (3)$$

$$c_{chlorophyll\text{-}b}(96\% \text{ ethanol}) = 24.96 \cdot A_{649} - 7.32 \cdot A_{665} \quad (4)$$

$$c_{chlorophyll\text{-}a}(100\% \text{ acetone}) = 11.75 \cdot A_{662} - 2.35 \cdot A_{645} \quad (5)$$

$$c_{chlorophyll\text{-}b}(100\% \text{ acetone}) = 18.61 \cdot A_{645} - 3.96 \cdot A_{662} \quad (6)$$

$$c_{chlorophyll\text{-}a}(95\% \text{ diethyl ether}) = 10.05 \cdot A_{662} - 0.76 \cdot A_{644} (7)$$

$$c_{chlorophyll\text{-}b}(95\% \text{ diethyl ether}) = 16.37 \cdot A_{644} - 3.14 \cdot A_{662}(8)$$

where A_λ is the absorbance at λ (in nm) wavelength, and $c_{chlorophyll\text{-}a}$ and $c_{chlorophyll\text{-}b}$ denote concentrations of chlorophyll *a* and chlorophyll *b* in µg cm^{-3} [12] as a function of solvents.

In Soxhlet extraction, the mass of the chlorophyll in the extract was calculated by the multiplication of the measured chlorophyll concentration values by the volume of the liquid in the round bottom flask. This also enabled the calculation of efficiency in mg of chlorophyll per g of dry algae units (relative to 100% dry algae powder).

In the leaching experiments, the chlorophyll concentration results calculated with the empirical formulae were multiplied by the volume of the solvent that resulted in the relative chlorophyll mass values in the given samples. Knowing the mass and water content of the algae, the efficiency of chlorophyll extraction relative to the mass of the dry algae can be calculated. These results were normalised to match the measurements and average values were calculated.
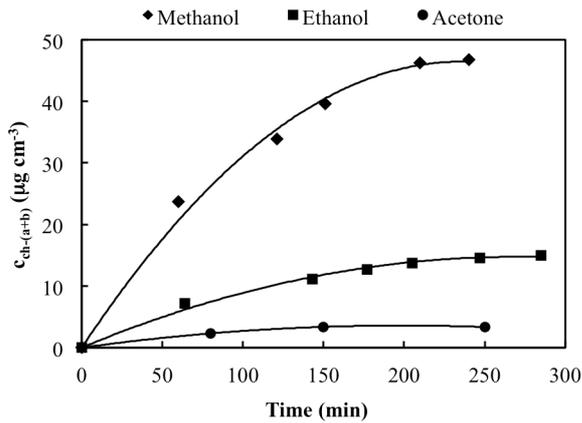
Figure 2: Changes in chlorophyll concentration during the Soxhlet measurements
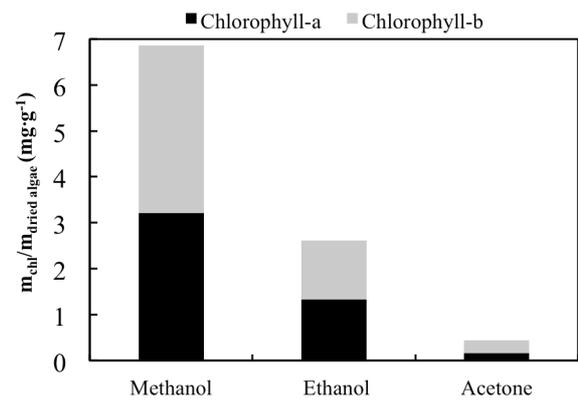
Figure 3: The efficiency of chlorophyll extraction of the various solvents during Soxhlet extraction
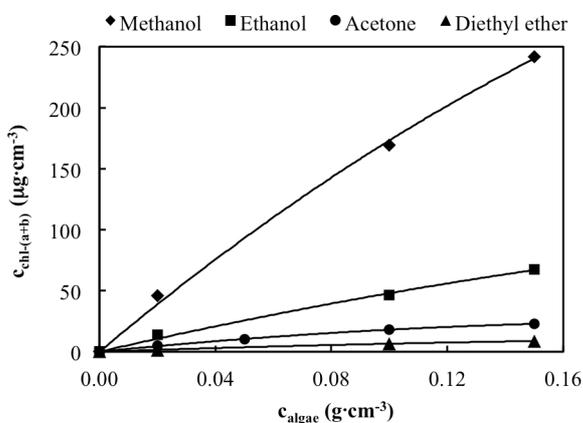
Figure 4: Change in chlorophyll concentration as a function of mass during leaching
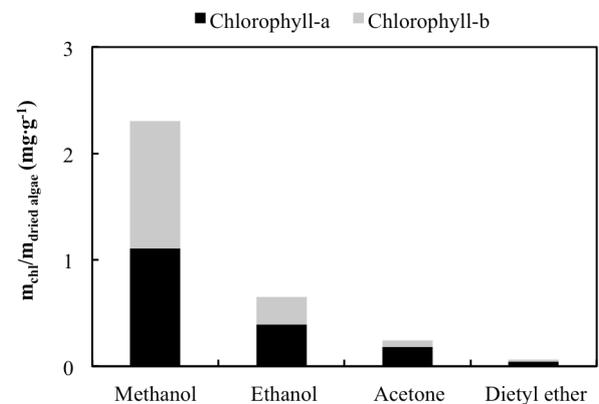
Figure 5: Efficiency of leaching using the various solvents

Results and Discussion

Soxhlet Extraction

The chlorophyll concentration measured changed according to a saturation curve as shown in *Fig.2*. It can be concluded that under the given circumstances the solvents reached their maximum efficiency and there was no need for longer extraction times. The ability of the solvents to extract chlorophyll relative to dried algae mass during the Soxhlet extraction is given in *Fig.3*.

The experiments were carried out using acetone, ethanol and methanol. It can be concluded that under the conditions of Soxhlet extraction, methanol is the most potent extracting solvent of chlorophyll *a* and *b* from the algae powder, followed by ethanol and acetone respectively. It can also be pointed out that acetone and methanol are more effective for the extraction of chlorophyll *b* while ethanol is more effective for the extraction of chlorophyll *a*. The latter experiment was not carried out with diethyl ether due to health and safety considerations.

Extraction by Leaching

Figs.4 and *5* illustrate the results of leaching. *Fig.4* depicts the concentration of chlorophyll for different masses of algae powder introduced into the system. The curves approach a saturation point in concentration, but do not reach the maximum. When we added progressively more and more algae powder to the same volume of solvent we observed increasing saturation behaviour. *Fig.5* presents the chlorophyll extracting ability of the various solvents when mixing 0.1 g of algae powder and 5 cm^3 of solvent. As expected, the best result was obtained by utilising at least 0.1 g of algae powder of. The efficiency ranking was similar to that of the Soxhlet extraction. Methanol proved to be the most efficient solvent, followed by ethanol, acetone and diethyl ether respectively.

The results are summarised in *Table 1,* which contains mg of chlorophyll per g of dry algae yields. This comparison enables the different solvents to be ranked and extraction methods, which is as follows according to their efficiency: methanol, ethanol, acetone, and diethyl ether. Additionally, significantly more chlorophyll can be extracted from ground algae with Soxhlet extraction than with leaching. In the

Table 1: Comparison of the chlorophyll (Chl.) extraction results

	Soxhlet extraction			Leaching		
	Chl. a	Chl. b	Chl. a & b	Chl. a	Chl. b	Chl. a & b
acetone	0.164	0.270	0.435	0.183	0.059	0.242
ethanol	1.329	1.290	2.619	0.395	0.259	0.654
methanol	3.206	3.657	6.862	1.109	1.199	2.307
diethyl ether	n/a	n/a	n/a	0.044	0.017	0.062

experiments conducted with acetone, it was observed that the extraction efficiency of chlorophyll a or chlorophyll b is greatly dependent on the conditions. Methanol, the solvent that proved to be the most efficient, could be removed later from the chlorophyll by low temperature evaporation.

Conclusions

From a systematic investigation of solid-phase extraction of chlorophyll a and chlorophyll b using Soxhlet extraction and leaching as a function of the employed solvent, it was found that the most efficient solvent is methanol. In both extraction techniques, the extracted amount of chlorophyll using methanol as a solvent is approximately an order of magnitude higher than in acetone and close to three to four times greater than the results of ethanol.

REFERENCES

[1] BOROWITZKA M.A.: Energy from Microalgae: A Short History, Develop. Appl. Phycology (Springer, The Netherlands), 2013, 5, 1-15

[2] NAJAFI G., GHOBADIAN B., YUSAF T.F.: Algae as a sustainable energy source for biofuel production in Iran: A case study, Renewable and Sustainable Energy Reviews, 2011, 15 (8), 3870–3876

[3] MATA T.M., MARTINS A.A., CAETANO N.S.: Microalgae for biodiesel production and other applications: A review, Renewable and Sustainable Energy Reviews, 2010, 14(1), 217–232

[4] BOCSI R., HORVATH G., HANAK L., HODAI Z.: Extraction examinations of microalgae propagated for biodiesel additives, Hung. J. Ind. Chem., 2011, 39(1), 45-49

[5] FAIZAL B.: Biotechnological applications of microalgae: biodiesel and value-added products, CRC Press, Bosa Roca, FL, United States, 2013, pp. 80

[6] DEMIRBAS A., DEMIRBAS M.F.: The importance of algae oil as a source of biodiesel, Energy Conversion and Management, 2011, 52(1), 163-170

[7] LANFER-MARQUEZ U.M., BARROS R.M.C., SINNECKER P.: Antioxidant activity of chlorophylls and their derivatives, Food Res. Int., 2005, 38(8-9), 885–891

[8] HOSIKIAN A., LIM S., HALIM R., DANQUAH M.K.: Chlorophyll extraction from microalgae: A review of the process engineering aspects, Int. J. Chem. Engng., 2010, Article ID 391632

[9] SIMPSON N.J.K., WELLS M.J.M.: Introduction to Solid-Phase Extraction, Marcel Dekker Inc, New York, NY, USA, 2000, pp. 1-18

[10] MEIRELES M.A.A.: Extracting Bioactive Compounds for Food Products, Theory and Applications Contemporary Food Engineering Series, CRC Press, Dublin, Ireland 2009, pp. 140-159

[11] LAN S., WU L., ZHANG D., HU C., LIU Y.: Ethanol outperforms multiple solvents in the extraction of chlorophyll-a from biological soil crusts, Soil Biology and Biochemistry, 2011, 43(4), 857–861

[12] WELLBURN A.R., LICHTENTHALER H.: Formulae and program to determine total carotenoids and chlorophylls-a and b of leaf extracts in different solvents, Adv. Agricul. Biotechn., 1984, 2(1), 9-12

CALCULATING THE ELECTROSTATIC POTENTIAL PROFILES OF DOUBLE LAYERS FROM SIMULATION ION DENSITY PROFILES

DEZSŐ BODA,[✉1] AND DIRK GILLESPIE[2]

[1]Department of Physical Chemistry, University of Pannonia, Egyetem u. 10., Veszprém, 8200, HUNGARY
[2]Department of Molecular Biophysics and Physiology, Rush University Medical Center, 1750 W. Harrison St., Chicago, USA
✉E-mail: boda@almos.vein.hu

Computer simulations of the planar double layer geometry provide the charge profile with statistical noise. To compute the mean electrostatic potential profile from the charge profile, one must solve Poisson's equation with appropriate boundary conditions (BC). In this work, we show that it is advantageous to use the Neumann or Dirichlet BCs at the boundaries of the simulation domain with an integrated version of Poisson's equation. This minimises errors from the simulation's noisy density profiles, in contrast to traditional convolution integrals that amplify the noise. The Neumann BC, where the electric field is prescribed, can be used in both the constant surface charge and constant electrode voltage ensembles. In the constant voltage ensemble, where the potential difference between the confining electrodes is prescribed, one can also use the Dirichlet BC, where the potentials at the boundaries are set. We show that the new methods provide converged results for the potential profile faster than the convolution integral does.

Keywords: electrical double layer, Poisson equation, boundary conditions, computer simulation

Introduction

The electrical double layer (DL) is formed by a charged surface and a phase (usually, a liquid) containing mobile charge carriers near the surface. Depending on the material carrying the mobile charges, DLs appear in electrolytes, molten salts, ionic liquids, plasmas, and even fast ion conductors (solid electrolytes). DLs in solutions of dissolved ions are particularly important in electrochemistry, biology, and colloid chemistry. DLs near electrodes differ from DLs near charged objects carrying a fixed surface charge (such as colloids, macromolecules, and porous bodies), because the surface charge on the electrode can be controlled by an external voltage.

Theoretical studies of DLs began with the Poisson-Boltzmann (PB) theory known as the Gouy-Chapman (GC) theory [1,2] in electrochemistry, the Debye-Hückel (DH) theory [3] in solution chemistry, and the Derjaguin-Landau-Verwey-Overbeek (DLVO) theory [4,5] in colloid chemistry. The PB theory is still very popular in applications because of its simplicity in spite of the fact that it neglects ionic correlations and effects due to the finite size of ions (e.g., excluded volume). More powerful statistical mechanical theories that are able to take these correlations into account have been developed [6–16]. Computer simulations are a versatile method of studying DLs in various geometries for various models of the constituents (ions, water, and electrode) [17–35]. This paper focuses on computing the electrical field and potential from a charge distribution obtained from computer simulations.

Computer simulations of systems containing charged particles must be performed in accordance with the laws of both electrostatics and statistical mechanics [36, 37]. This means that the electrical potential must be computed accurately for every configuration sampled in a Monte Carlo (MC) simulation. Alternatively, the electrical field must be computed accurately in every time step of a molecular dynamics (MD) simulation. Every configuration of the system (positions of ions) corresponds to a microscopic state. In this case, we consider the system at the microscopic level. A simulation that samples the possible microscopic states must be performed properly according to the probability distribution of a given statistical mechanical ensemble. One typical simulation method to handle various ensembles is MC. The density distribution of ionic species i, $\langle \rho_i(\mathbf{r}) \rangle$, is obtained as an ensemble average from the simulation. From these, the average charge distribution can be obtained as

$$q(\mathbf{r}) = \langle Q(\mathbf{r}) \rangle = \sum_i z_i e \langle \rho_i(\mathbf{r}) \rangle, \qquad (1)$$

where $Q(\mathbf{r})$ is the charge distribution in a microscopic state, e is the electronic unit charge and z_i is the valence of ionic species i.

The average (mean) electrostatic potential can be obtained "on the fly" by computing the potential in the simulation cell for every configuration (denoted by $\Psi(\mathbf{r})$) and then taking the ensemble average, $\psi(\mathbf{r}) = \langle \Psi(\mathbf{r}) \rangle$. In this work, we use upper-case symbols for the microscopic quantities ($\Psi(\mathbf{r})$ and $Q(\mathbf{r})$), while we use lower-case symbols for their macroscopic counterparts, namely,

their ensemble averages ($\psi(\mathbf{r})$ and $q(\mathbf{r})$). The electrical potential in a configuration of the MC simulation, $\Psi(\mathbf{r})$, can be computed analytically from Coulomb's law or numerically using a Poisson solver. In both cases, we apply electrostatic BCs at the microscopic level.

The reverse order, when we compute the ensemble average of the charge distribution *(Eq.(1))* and then solve Poisson's equation for the mean potential

$$\nabla^2 \psi(\mathbf{r}) = -\frac{1}{\epsilon_0} q(\mathbf{r}) \qquad (2)$$

is more usual (ϵ_0 is the permittivity of vacuum). This equation applies to either explicit or implicit solvent models. In the explicit solvent framework, $q(\mathbf{r})$ also contains the charge distribution of the water molecules (in addition to their ionic charges). In the implicit solvent framework, where water is represented by a dielectric background characterised by a dielectric constant, ϵ, $q(\mathbf{r})$ also contains polarisation charges induced in the dielectric (in the simplest case, $q(\mathbf{r})$ is the ionic charges divided by ϵ). Because we are solving a differential equation, we must use appropriate boundary conditions (BCs) when solving *Eq.(2)*. Because we want an ensemble averaged result, we apply the BCs at the macroscopic level.

This paper describes how to numerically integrate the Poisson equation with appropriate BCs to efficiently compute the mean electrostatic potential profile from the mean charge profile obtained from simulations. The statistical ensemble applied in the simulation determines which BC to use. In this paper, we consider the two basic ensembles, one where the electrode charges are fixed (constant charge ensemble) and one where the difference between the electrode potentials is fixed (constant voltage ensemble).

In the constant charge ensemble, the traditional method of computing the potential profile is via a convolution integral. Here, we show that this method is error-prone and numerically inefficient and instead, it is more advantageous to use a different integration scheme with Neumann BCs, where the normal electric fields at the boundaries of the system are fixed [38].

In the constant voltage ensemble developed by KIY-OHARA and ASAKA [39], the electrode potentials are known in advance. Therefore, we can also use Dirichlet BCs, where the potentials at the boundaries of the system are prescribed.

In the following, we describe our model in detail. Then, we consider all three issues discussed above (BCs at the microscopic level, statistical mechanical ensembles, and BCs at the macroscopic level) and present various possibilities for the macroscopic BC depending on the statistical mechanical ensemble used. We present results of model calculations to show the self-consistency of these calculations.

Model and Boundary Condition at the Microscopic Level

At the microscopic level, we have a system that contains localised discrete and/or continuous distributions of charges $Q(\mathbf{r})$. In practice, there are two traditional schools to compute the electrostatic energy (in MC) or forces (in MD) for a configuration sampled by a computer simulation. In one school, the potential is computed on a grid from Poisson's equation *(Eq.(2))* using a partial differential equation (PDE) solver with appropriate boundary conditions. This method is generally used in MD simulations of explicit solvent systems.

In the other school, Coulomb's law is used to calculate the potential:

$$\Psi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0\epsilon} \int Q(\mathbf{r}')G(\mathbf{r}, \mathbf{r}')d\mathbf{r}', \qquad (3)$$

where $G(\mathbf{r}, \mathbf{r}')$ is the appropriate Green's function. This method is generally used for simulations with implicit solvents.

The planar DL geometry means that we have a rectangular simulation cell of length $L_2 - L_1$ and with a base $H \times H$ confined by two planar electrodes at the two ends ($x = L_1$ and $x = L_2$) carrying surface charges σ_1 and σ_2. Periodic boundary conditions (PBC) are used in the y, and z dimensions, which means that the Green's function is

$$G(\mathbf{r}, \mathbf{r}') = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \frac{1}{|\mathbf{r} - \mathbf{r}' + jH\mathbf{n}_y + kH\mathbf{n}_z|} \quad (4)$$

where the sum over j and k represents the interaction with the periodic image charges in the replicas of the central simulation cell in the y and z dimensions set by the unit vectors $\mathbf{n}_y$ and $\mathbf{n}_z$. The term $j = k = 0$ corresponds to the interaction with the charge in the central simulation cell computed explicitly. In this paper, the interaction with periodic replicas is taken into account using the charged sheet method by smearing these charges into a sheet carrying q/H^2 surface charges with a H^2 square hole in the middle. The interaction with the holed sheet can be integrated. For further details, see Ref. [18, 29].

The statistical mechanical ensemble determines which thermodynamic variables are fixed in the simulation. The attempts in the MC simulations are designed to ensure sampling according to the probability distribution of the given ensemble. Moreover, the simulation must be self-consistent in the sense that a prescribed thermodynamic variable must agree with its value computed as an output of the simulation. For example, in an NpT simulation (where the pressure is fixed), the pressure can also be computed from the virial sum as an ensemble average in both NVT and NpT simulations. This value must be equal to the one prescribed in the NpT ensemble. In practice, however, the accuracy of the pressure computed from the virial sum depends on the size of the system, and one obtains the same relationship between pressure

and density from NVT and NpT simulations only in the limit of very large simulation cells.

From the point of view of the DL problem, the question is whether we perform the simulation in the constant charge or constant voltage ensemble. In the constant charge ensemble the surface charges on the confining walls of the simulation cell are fixed. This is the traditional simulation setup used in DL simulations since TORRIE and VALLEAU [17, 18]. The output of the simulation is the density profiles of the various ionic species $\langle \rho_i(x) \rangle$ (the PBC applied in the y and z dimensions ensures that the profiles depend on x only), from which the x-dependent charge profile, $q(x)$, is obtained (see *Eq.(1)*). The corresponding form of Poisson's equation is

$$\frac{d^2\psi(x)}{dx^2} = -\frac{1}{\epsilon_0\epsilon}q(x), \qquad (5)$$

where ϵ is now included so $q(x)$ constrains only ionic and electrode charges.

Recently, KIYOHARA and ASAKA [39] introduced the constant voltage ensemble, where the potential difference between the two confining walls (the electrodes) is prescribed. A special MC step, where a small amount of charge, $\Delta\sigma$, is transferred from one electrode to the other, has been introduced. The charge exchange is accepted on the basis of the Boltzmann factor

$$\min\left[1, \exp\left(-\frac{\Delta U}{kT} + \frac{H^2\Delta\psi^*\Delta\sigma}{kT}\right)\right] \qquad (6)$$

where ΔU is the energy change associated with the charge movement and $\Delta\psi^* = \psi(L_2) - \psi(L_2)$ is the prescribed voltage. Here, the surface charges on the electrodes fluctuate, while the potential difference between the electrodes ($\Delta\psi^*$) is an independent, prescribed variable of the ensemble. The mean potential profile can be computed from *Eq.(5)*. The computed voltage, namely, the potential difference between the electrodes, $\Delta\psi = \psi(L_2) - \psi(L_2)$, must be equal to the prescribed voltage, $\Delta\psi^*$.

In the following, we describe various ways to apply BCs at the macroscopic level using a case study of a 1:1 electrolyte, where the ions are modeled as charged hard spheres with diameters $d_+ = d_- = 3$ Å and the dielectric constant of water is 78.46. The concentration is 1 M and the temperature is 298.2 K. We show results for the special case of $\sigma_1 = -\sigma_2 = \sigma = 0.1$ Cm^{-2} and $L_1 = -L_2 = L$, but the equations are presented for the general case. We used 200/200 ions in the MC simulations performed in the canonical (NVT) ensemble. The dimensions of the cell are $L = 50$ Å and $H = 57.2$ Å.

Boundary Conditions at the Macroscopic Level

Boundary Conditions Set in the Bath: The Convolution Integral

Let us distinguish between the charge of the ions obtained from the simulation, $q_{\text{ion}}(x)$, and the surface charges,

$\sum_k \sigma_k \delta(x - x_k)$, at $x_1 = L_1$ and $x_2 = L_2$ (in this work, we assume only two charged surfaces at L_1 and L_2, but there can be more as in the simulations of KIYOHARA *et al.* [40–50] for porous electrodes). The total charge is the sum of these:

$$q(x) = q_{\text{ion}}(x) + \sum_k \sigma_k \delta(x - x_k). \qquad (7)$$

The ionic charge profile is obtained as an ensemble average. The electrode charges are prescribed in the constant charge ensemble, while they are obtained as ensemble averages in the constant voltage ensemble.

At the macroscopic level, the issue of electrostatic self-consistency appears when we ask the question: what kind of BC should be applied when we solve Poisson's equation *(Eq.(5))*. The traditional answer to this question is that the BC is set in the bulk electrolyte, where the average electrical field and electrical potential are zero. The corresponding solution of Poisson's equation can then be obtained in the form of a convolution integral

$$\psi(x) = -\frac{1}{\epsilon_0\epsilon}\int_x^\infty (x' - x)q_{\text{ion}}(x')dx'. \qquad (8)$$

This solution was probably inspired by theories that usually consider an isolated DL where the BCs are set in infinity. Because theories (unlike simulations) provide smooth charge profiles (with the property $\lim_{x\to\infty} q(x) = 0$) without any noise, this integral works well for theories. To the best of our knowledge, the majority of researchers (among others, the authors of this paper) have used this equation in the past [6–35]. In this work, we show that this equation, from a numerical point of view, is a poor choice to compute the potential in simulation studies. The reason is that the upper integration limit is not well defined and that $q(x)$ is subject to a large statistical noise.

In practice, the upper integration limit is set somewhere in the middle of the cell where a bulk electrolyte is. Here we will use exactly the middle of the cell ($x = 0$) as the upper integration limit. Because of $x - x'$ in the integrand, this integral is very sensitive to the noise in $q(x)$, because the noise is amplified as one moves further away from the electrode.

The results of a very short (200 MC cycles; 1000 attempts to move ions were made in an MC cycle) simulation are shown in *Fig.1*. The density and charge profiles are very noisy. When we compute the potential from *Eq.(8)*, the result is subject to a large error and is far from what we are supposed to get (*Fig.2*, top panel). The slope is not necessarily zero in the bath (which means that the electrical field is not zero). If the simulation was run for longer, we would get a completely different result. In general, long simulations are needed to produce a smooth charge profile and a well established potential profile. The problem is even more serious when we try to reproduce small effects, such as the value of the electrode potential at zero electrode charge (PZC) for asymmetric electrolytes. The value of the PZC potential is very small
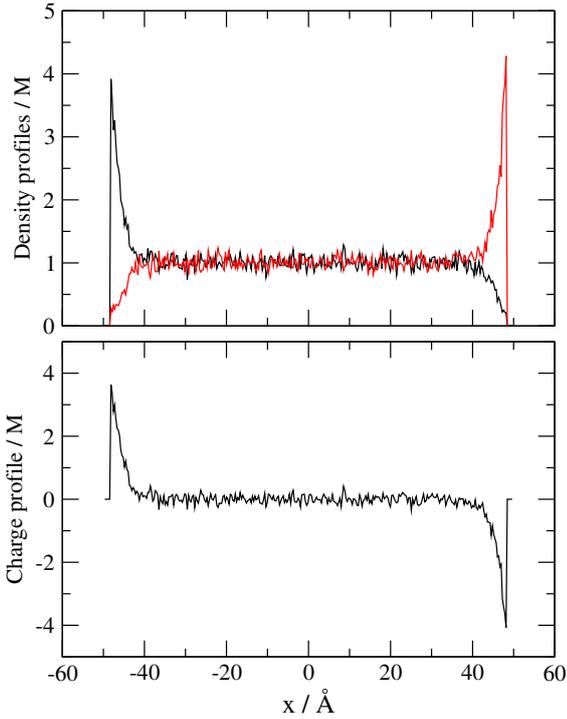
Figure 1: Density profiles (top panel) and a charge profile (bottom panel) obtained from a short (200 MC cycles) constant charge simulation using $\sigma_1 = -\sigma_2 = -0.1$ Cm^{-2}. The charge profile is obtained from $\sum_i z_i \rho_i(x)$, so its unit is M (mol dm^{-3})

and the effect of the noise is dramatic. Extremely long simulations are needed to obtain convergent results for the potential [51].

To illustrate this weak convergence, we have plotted the left electrode potential ($\psi(L_1) - \psi(0)$, top panel) and the voltage ($\psi(L_2) - \psi(L_1)$, bottom panel) as computed from *Eq.(8)* as functions of the performed MC cycles in *Fig.3* (red dashed curves). These potential values fluctuate strongly. The calculated values depend not only on the simulation time, but on the upper integration limit. If we shift that point a bit, we get a different result (data not shown).

Neumann Boundary Conditions and the Constant Charge Ensemble

Here we propose, instead, to use Neumann or Dirichlet BCs at the boundaries of the simulation cell (at L_1 and L_2, or, equivalently, at $-\infty$ and ∞). In the case of the Neumann BC, the normal electrical field is prescribed, while in the case of the Dirichlet BC, the electrical potential is prescribed at the confining walls. Our simulation setup ensures that the simulation cell is always charge neutral. Then, Gauss law states that the average electrical field is zero outside the cell for the regions $z < L_1$ and $z > L_2$. This information makes the Neumann BC applicable both in the constant charge and constant voltage
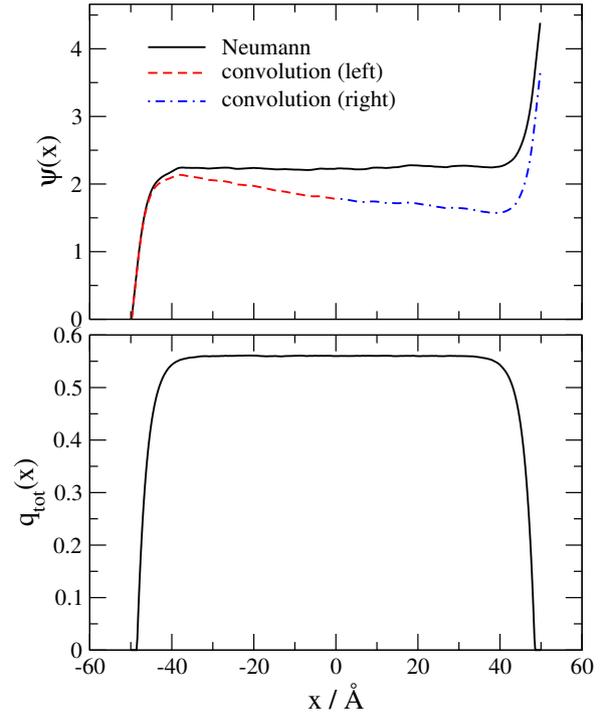
Figure 2: Potential profiles (top panel) as computed from the Neumann BC *(Eq.(11))* and the convolution integral *(Eq.(8))*. The upper limit of the convolution integral is $x = 0$ and it is computed for the left- and right-hand sides separately. The bottom panel shows the integral of the charge profile ($q_{\mathrm{tot}}(x)$, see *Eq.(7)*). It is closely related to the electric field through *Eq.(13)*. The profiles have been obtained from the curves of *Fig.1*. The electrostatic potential is shown in units of kT/e throughout this paper

ensembles.

In the constant charge ensemble, the potential difference between the electrodes is an output of the calculation. Therefore, we cannot use it as a BC, so we cannot apply the Dirichlet BC in this case. In the constant voltage ensemble, on the other hand, the potential difference is known in advance (see the next section). In the constant voltage ensemble, therefore, both Neumann and Dirichlet BCs can be used and they should give the same answer (apart from errors related to the size of the system, see later).

The Neumann BC for *Eq.(8)* is that the electrical field is zero outside the system:

$$E(x \to -\infty) = - \left. \frac{d\psi(x)}{dx} \right|_{x \to -\infty} = 0 \qquad (9)$$

and the same for $x \to \infty$. By integrating Poisson's equation once, we obtain

$$\frac{d\psi(x)}{dx} = -\frac{1}{\epsilon_0 \epsilon} \int_{-\infty}^{x} q(x') dx' + C_1 \qquad (10)$$

where C_1 is an integration constant. Taking *Eq.(10)* at
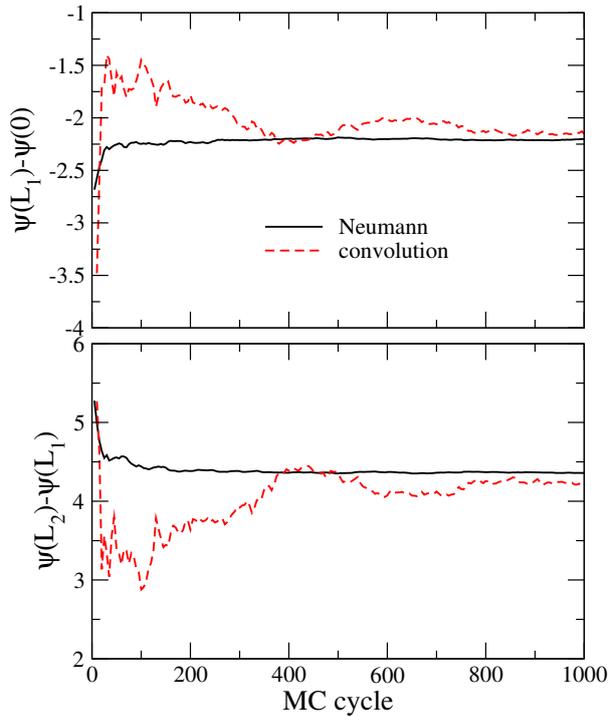
Figure 3: Convergence of the left electrode potential (top panel) and the voltage (bottom panel) as computed from the Neumann BC *(Eq.(11))* and the convolution integral *(Eq.(8))* in the constant charge ensemble.

Figure 4: Convergence of the left electrode potential (top panel) and the voltage (bottom panel) as computed from the Neumann BC *(Eq.(11))*, the Dirichlet BC *(Eq.(17))*, and the convolution integral *(Eq.(8))* in the constant voltage ensemble

any location $x < L_1$ (where $q(x) = 0$) and using the BC *(Eq.(9))*, we get $C_1 = 0$ for the integration constant. If we use *Eq.(7)* for $q(x)$, we obtain

$$\frac{d\psi(x)}{dx} = -\frac{1}{\epsilon_0 \epsilon} \int_{L_1}^{x} q_{\mathrm{ion}}(x') dx' - \frac{1}{\epsilon_0 \epsilon} \sigma_1 \qquad (11)$$

for $L_1 < x < L_2$. The change in the lower integration limit was possible because ions exist only between the two electrodes ($q_{\mathrm{ion}} \neq 0$ only for $L_1 < x < L_2$). Introducing

$$q_{\mathrm{tot}}(x) = \int_{L_1}^{x} q_{\mathrm{ion}}(x') dx' \qquad (12)$$

for the integral of the ionic charge profile (the total charge density per area in the $[L_1, x]$ interval), the electrical field can be given as

$$E(x) = \frac{1}{\epsilon_0 \epsilon} q_{\mathrm{tot}}(x) + \frac{1}{\epsilon_0 \epsilon} \sigma_1 \qquad (13)$$

for $L_1 < x < L_2$. The $q_{\mathrm{tot}}(x)$ profile is shown in the bottom panel of *Fig.2*. For the special case of $\sigma_1 = -\sigma_2$, the total ionic charge is zero in the system, so $q_{\mathrm{tot}}(L_2) = 0$. The nearly constant but noisy profile in the middle of the cell represents the bulk region.

Integrating once more, we obtain

$$\psi(x) = -\frac{1}{\epsilon_0 \epsilon} \int_{L_1}^{x} q_{\mathrm{tot}}(x') dx' - \frac{1}{\epsilon_0 \epsilon} \sigma_1 (x - L_1) + C_2, \qquad (14)$$

where C_2 is another integration constant. Its value is inconsequential because we can set the zero level of the potential arbitrarily. *Fig.2* shows the $\psi(x)$ profile where $C_2 = 0$. It is usual, however, to set the ground in the bulk, so $C_2 = -\overline{\psi(0)}$, where $\overline{\psi(0)}$ is the average of the potential profile over the bulk. KIYOHARA and ASAKA [39, 52] used an equation of a convolution form that can be shown to be equivalent to *Eq.(14)*, but they seemed to leave out the linear term containing σ_1.

Eq.(14) is different from the convolution integral *(Eq.(8))* because it does not suffer from the uncertainty in the upper integration limit and from errors originating from the noise of the $q_{\mathrm{ion}}(x)$ profile. The integration is performed for a well defined finite domain from the left electrode to x. *Fig.3* shows that this equation provides a much better convergence as a function of simulation time for the same simulation (same $q_{\mathrm{ion}}(x)$).

This procedure was used in our papers for inhomogeneous electrolyte systems to compute the mean electrostatic potential [53–56]. It was especially useful for electrolytes adsorbed in narrow slits [55, 56]. In this case, the DLs formed at the walls of the slit overlap so a bulk electrolyte does not form in the middle of the slit and the slit is not charge neutral. The Neumann BC is then the natural BC so the electric field is zero behind the walls. In another example, the Neumann BC is used to compute the potential for a DL model, where the electrode, the

inner layer, and the electrolyte have different dielectric constants [54]. In this case, the polarisation charge induced at the dielectric boundaries must also be included in *Eq.(14)*.

Dirichlet Boundary Conditions and the Constant Voltage Ensemble

An alternative method is the constant voltage ensemble of KIYOHARA and ASAKA [39]. Here, the electrode charges fluctuate and the potential between the electrodes is prescribed. Therefore, we can also apply Dirichlet BCs, where

$$\psi(L_1) = 0 \qquad (15)$$

and

$$\psi(L_2) = \Delta\psi^*. \qquad (16)$$

The general solution for the potential profile in the $L_1 < x < L_2$ range is

$$\psi(x) = -\frac{1}{\epsilon_0\epsilon} \int_{L_1}^{x} q_{\text{tot}}(x')dx' + C_1(x - L_1) + C_2, \quad (17)$$

where we integrate from the right-hand side of the electrode at $x = L_1$, so the surface charge σ_1 is now excluded from the integration. The BC at $x = L_1$ *(Eq.(15))* provides the integration constant $C_2 = 0$, while the BC at $x = L_2$ *(Eq.(16))* provides the integration constant

$$C_1 = \frac{1}{L_2 - L_1} \left[\Delta\psi^* + \frac{1}{\epsilon_0\epsilon} \int_{L_1}^{L_2} q_{\text{tot}}(x)dx \right]. \quad (18)$$

Of course, we can calculate the potential profile using the Neumann BCs too. In that case, we must use $\langle\sigma_1\rangle$ in *Eq.(14)* instead of σ_1 because the electrode charge is now fluctuating so its value is not known in advance. Therefore, its ensemble average should be used in *Eqs.(7-14)*.

In the constant voltage ensemble we need the value of the voltage that corresponds to $\sigma_1 = -0.1$ Cm^{-2} as used in the previous constant charge simulation. This value we estimated with a very long (50,000 MC cycles) constant charge simulation and was obtained as $\Delta\psi^* = 4.398\,kT/e$. This value was used in the constant voltage simulation as an input parameter.

Fig.4 is the analogous version of *Fig.3*. The black solid curves are the results obtained from the Dirichlet BCs *(Eqs.(17)* and *(18))*. The red short-dashed curves show the results of the Neumann BC using $\langle\sigma_1\rangle$ in *Eq.(14)*. The convolution integral results (blue long-dashed curves) are inaccurate and poorly converged for such a short simulation.

In the constant voltage ensemble the electrode charge is a fluctuating quantity. *Fig.5* shows its convergence. Its limiting value is not equal to that used in the constant charge simulation that provided the input voltage value $\Delta\psi^* = 4.398\,kT/e$. The deviation is due to finite size of the system. Using a larger simulation box (larger H), a smaller deviation is observed (data not shown). This deviation is also observed in *Fig.4*. The limiting value of the
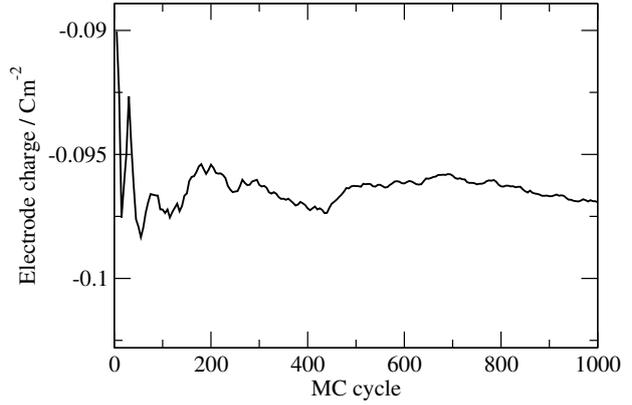
Figure 5: Convergence of the electrode charge in the constant voltage ensemble

potential difference $\Delta\psi = \psi(L_2) - \psi(L_1)$ obtained from the calculation using the Neumann BC (bottom panel) is different from the prescribed value (the value used in the Dirichlet BC, see also bottom panel).

This behaviour is analogous to that discussed earlier regarding the example of the pressure computed in the NVT and NpT ensembles. The constant charge ensemble corresponds to the NVT, while the constant voltage ensemble corresponds to the NpT ensemble. In the constant charge ensemble, the voltage is computed using the Neumann BC, just as the pressure is computed in the NVT ensemble from the virial sum. In the constant voltage ensemble, the voltage is prescribed, just as the pressure is prescribed in the NpT ensemble. The voltage can also be computed from the Neumann BC, just as the pressure can also be computed from the virial sum in the NpT ensemble.

Conclusion

We propose that the Neumann or Dirichlet BCs should be used in computing the mean electrostatic potential for the planar DL geometry studied by computer simulations. The commonly used convolution integral of *Eq.(8)* requires a vaguely defined upper integration limit and also suffers from numerical problems because it magnifies the effect of the noise in the charge profile that is always present in computer simulations. The problem of noise in the density profiles is unique to simulations and therefore more care must be taken in computing the electrostatic potential. On the other hand, theories [6–16], which produce smooth, noise-free density profiles, can use the convolution integral.

We have shown here that the numerical method we proposed is much more efficient than the convolution integral, because we use unambiguous parameters in the BC, specifically, the electrode charge in the case of the Neumann BC and the voltage in the case of the Dirichlet BC. Also, the simulation cell is necessarily finite, therefore, the boundaries of the system are always well de-

fined. Overall, our method leads to converged results with very short simulations.

Acknowledgement

REFERENCES

[1] GOUY G.: Sur la constitution de la charge electrique a la surface d'un electrolyte (English Title Please), J. Phys. (Paris), *1910*, 9, 457 (in French)

[2] CHAPMAN D.L.: A Contribution to the Theory of Electrocapillarity, Phil. Mag., *1913*, 25, 475

[3] DEBYE P., HÜCKEL E.: The theory of electrolytes. I. Lowering of freezing point and related phenomena, Physik. Z., *1923*, 24, 185–206

[4] DERJAGUIN B., LANDAU L.: Theory of the stability of strongly charged lyophobic sols and of the adhesion of strongly charged particles in the solution of electrolytes, Acta Physiochem USSR, *1941*, 14, 633–662

[5] VERWEY E.J.W., OVERBEEK J.T.G.: Theory of the Stability of Lyophobic Colloids Elsevier, Amsterdam, *1948*

[6] BLUM L.: Theory of electrified interfaces, J. Phys. Chem., *1977*, 81(2), 136–147

[7] HENDERSON D., BLUM L., SMITH W.R.: Application of the hypernetted chain approximation to the electric double-layer at a charged planar interface, Chem. Phys. Lett., *1979*, 63(2), 381–383

[8] BLUM L., HENDERSON D.: Mixtures of hard ions and dipoles against a charged wall: The Ornstein–Zernike equation, some exact results, and the mean spherical approximation, J. Chem. Phys., *1981*, 74(3), 1902–1910

[9] OUTHWAITE C.W., BHUIYAN L.B.: An improved modified Poisson-Boltzmann equation in electric-double-layer theory, J. Chem. Soc. Faraday. Trans. II., *1983*, 79, 707–718

[10] LOZADA-CASSOU M., HENDERSON D.: Application of the hypernetted chain approximation to the electrical double-layer - comparison with Monte Carlo results for 2-1 and 1-2 salts, J. Phys. Chem., *1983*, 87(15), 2821–2824

[11] MIER-Y TERAN L., SUH S.H., WHITE H.S., DAVIS H.T.: A nonlocal free-energy density-functional approximation for the electrical double-layer, J. Chem. Phys., *1990*, 92(8), 5087–5098

[12] KIERLIK E., ROSINBERG M.L.: Free-energy density functional for the inhomogeneous hard-sphere fluid - application to interfacial adsorption, Phys. Rev. A, *1990*, 42(6), 3382–3387

[13] ROSENFELD Y.: Free-energy model for inhomogeneous fluid mixtures - Yukawa-charged hard-spheres, general interactions, and plasmas, J. Chem. Phys., *1993*, 98(10), 8126–8148

[14] GILLESPIE D., NONNER W., EISENBERG R.S.: Density functional theory of charged, hard-sphere fluids, Phys. Rev. E, *2003*, 68(3), 031503

[15] DI CAPRIO D., STAFIEJ J., BADIALI J.P.: Field theoretical approach to inhomogeneous ionic systems: thermodynamic consistency with the contact theorem, Gibbs adsorption and surface tension, Mol. Phys., *2003*, 101(16), 2545–2558

[16] PIZIO O., PATRYKIEJEW A., SOKOLOWSKI S.: Phase behavior of ionic fluids in slitlike pores: A density functional approach for the restricted primitive model, J. Chem. Phys., *2004*, 121(23), 11957–11964

[17] TORRIE G.M., VALLEAU J.P.: Monte Carlo study of an electrical double-layer, Chem. Phys. Lett., *1979*, 65(2), 343–346

[18] TORRIE G.M., VALLEAU J.P.: Electrical double-layers 1. Monte Carlo study of a uniformly charged surface, J. Chem. Phys., *1980*, 73(11), 5807–5816

[19] TORRIE G.M., VALLEAU J.P., PATEY, G.N.: Electrical double-layers 2. Monte Carlo and HNC studies of image effects, J. Chem. Phys., *1982*, 76(9), 4615–4622

[20] VALLEAU J.P., TORRIE G.M.: The electrical double-layer 3. Modified Gouy-Chapman theory with unequal ion sizes, J. Chem. Phys., *1982*, 76(9), 4623–4630

[21] TORRIE G.M., VALLEAU J.P.: Electrical double-layers 4. Limitations of the Gouy-Chapman theory, J. Phys. Chem., *1982*, 86(16), 3251–3257

[22] VALLEAU J.P., TORRIE G.M.: Electrical double-layers 5. Asymmetric ion wall interactions, J. Chem. Phys., *1984*, 81(12), 6291–6295

[23] TORRIE G.M., VALLEAU J.P., OUTHWAITE C.W.: Electrical double-layers 6. Image effects for divalent ions, J. Chem. Phys., *1984*, 81(12), 6296–6300

[24] VAN MEGEN W., SNOOK I.: The grand canonical ensemble Monte Carlo method applied to the electrical double-layer, J. Chem. Phys., *1980*, 73(9), 4656–4662

[25] SNOOK I., VAN MEGEN W.: Finite ion size effects in the electrical double-layer - a Monte Carlo study, J. Chem. Phys., *1981*, 75(8), 4104–4106

[26] BRATKO D., JÖNSSON B., WENNERSTRÖM H.: Electrical double layer interactions with image charges, Chem. Phys. Lett., *1986*, 128, 449–454

[27] TANG Z.X., SCRIVEN L.E., DAVIS H.T.: A 3-component model of the electrical double-layer, J. Chem. Phys., *1992*, 97(1), 494–503

[28] PHILPOTT M.R., GLOSLI J.N.: Molecular dynamics simulation of interfacial electrochemical processes: Electric double layer screening, Solid-liquid Electrochemical Interfaces, *1997*, 656, 13–30

[29] BODA D., CHAN K.Y., HENDERSON D.: Monte Carlo simulation of an ion-dipole mixture as a model of an electrical double layer, J. Chem. Phys., *1998*, 109(17), 7362–7371

[30] SPOHR E.: Computer simulation of the structure of the electrochemical double layer, J. Electroanalytical Chem., *1998*, 450(2), 327–334

[31] BODA D., HENDERSON D., CHAN K.Y.: Monte Carlo study of the capacitance of the double layer in a model molten salt, J. Chem. Phys., *1999*, 110(11), 5346–5350

[32] CROZIER P.S., ROWLEY R.L., HENDERSON D.: Molecular-dynamics simulations of ion size effects on the fluid structure of aqueous electrolyte systems between charged model electrodes, J. Chem. Phys., *2001*, 114(17), 7513–7517

[33] GUYMON C.G., HUNSAKER M.L., HARB J.N., HENDERSON D., ROWLEY R.L.: Effects of solvent model flexibility on aqueous electrolyte behavior between electrodes, J. Chem. Phys., *2003*, 118(22), 10195–10202

[34] SPOHR E.: Some recent trends in computer simulations of aqueous double layers, Electrochimica Acta, *2003*, 49(1), 23–27

[35] HENDERSON D., BODA D.: Insights from theory and simulation on the electrical double layer, Phys. Chem. Chem. Phys., *2009*, 11(20), 3822–3830

[36] ALLEN M.P., TILDESLEY D.J.: Computer Simulation of Liquids Oxford, New York, *1987*

[37] FRENKEL D., SMIT B.: Understanding molecular simulations Academic Press, San Diego, *1996*

[38] JACKSON J.D.: Classical Electrodynamics Wiley, New York, 3rd edn., *1999*

[39] KIYOHARA K., ASAKA K.: Monte Carlo simulation of electrolytes in the constant voltage ensemble, J. Chem. Phys., *2007*, 126(21), 214704 (pages 14)

[40] KIYOHARA K., ASAKA K.: Monte Carlo simulation of electrolytes in the constant voltage ensemble, J. Chem. Phys., *2007*, 126(21), 214704

[41] KIYOHARA, K., ASAKA, K.: Monte Carlo simulation of porous electrodes in the constant voltage ensemble, J. Phys. Chem. C, *2007*, 111(43), 5903–15909

[42] KIYOHARA K., SUGINO T., TAKEUCHI I., MUKAI K., ASAKA K.: Expansion and contraction of polymer electrodes under an applied voltage, J. Appl. Phys., *2009*, 105(6), 063506

[43] KIYOHARA K., SUGINO T., TAKEUCHI I., MUKAI K., ASAKA K.: Expansion and contraction of polymer electrodes under an applied voltage, J. Appl. Phys., *2009*, 105(11), 119902

[44] KIYOHARA K., SUGINO T., ASAKA K.: Electrolytes in porous electrodes: Effects of the pore size and the dielectric constant of the medium, J. Chem. Phys., *2010*, 132(14), 144705

[45] KIYOHARA K., SUGINO T., ASAKA K.: Phase transition in porous electrodes, J. Chem. Phys., *2011*, 134(15), 154710

[46] KIYOHARA K., SUGINO T., ASAKA K.: Molecular mechanism of ionic electroactive polymer actuators, Smart Mat. & Struc., *2011*, 20(12), 124009

[47] KIYOHARA K., SHIOYAMA H., SUGINO T., ASAKA K.: Phase transition in porous electrodes. II. Effect of asymmetry on ion size, J. Chem. Phys., *2012*, 136(9), 094701

[48] KIYOHARA K., ASAKA K.: Voltage induced pressure in porous electrodes, Mol. Phys., *2013*, 111(2), 295–306

[49] KIYOHARA K., SHIOYAMA H., SUGINO T., ASAKA K., SONEDA Y., IMOTO K., KODAMA M.: Phase transition in porous electrodes. III. In the case of a two component electrolyte, J. Chem. Phys., *2013*, 138(23), 234704

[50] KIYOHARA K., SHIOYAMA H., SUGINO T., ASAKA K., SONEDA Y., IMOTO K., KODAMA M.: Phase transition in porous electrodes. III. In the case of a two component electrolyte (vol 138, 234704, 2013), J. Chem. Phys., *2013*, 139(6), 069902

[51] VALISKÓ M., HENDERSON D., BODA D.: Competition between the effects of asymmetries in ion diameters and charges in an electrical double layer studied by Monte Carlo simulations and theories, J. Phys. Chem. B, *2004*, 108(42), 16548–16555

[52] KIYOHARA K., ASAKA K.: Monte Carlo Simulation of Porous Electrodes in the Constant Voltage Ensemble, J. Phys. Chem. C, *2007*, 111(43), 15903–15909

[53] NAGY T., VALISKÓ M., HENDERSON D., BODA D.: The Behavior of 2:1 and 3:1 Electrolytes at Polarizable Interfaces, J. Chem. Eng. Data, *2011*, 56(4), 1316–1322

[54] NAGY T., HENDERSON D., BODA D.: Simulation of an electrical double layer model with a low dielectric layer between the electrode and the electrolyte, J. Phys. Chem. B, *2011*, 115(39), 11409–11419

[55] KOVÁCS R., VALISKÓ M., BODA D.: Monte Carlo simulation of the electrical properties of electrolytes adsorbed in charged slit-systems, Cond. Matt. Phys., *2012*, 15(2), 23803

[56] VALISKÓ M., HENDERSON D., BODA D.: Selective adsorption of ions in charged slit systems, Cond. Matt. Phys., *2013*, 16(4), 43601

DEGRADATION OF REINFORCED AND UNREINFORCED WASTE POLYAMIDES DURING MECHANICAL RECYCLING

János Sója and Norbert Miskolczi[✉]

MOL Department of Hydrocarbon and Coal Processing, Institute of Chemical Engineering and Process Engineering, University of Pannonia, Egyetem u. 10, 8200, HUNGARY
[✉]Email: mnorbert@almos.uni-pannon.hu

The paper discusses the degradation of waste polyamides during mechanical recycling from the automotive sector. Two different polyamides were investigated: glass fibre reinforced and unreinforced. Raw materials were reprocessed twenty times and the changes in their properties were investigated as a function of reprocessing number. Considerable differences were found in relation to the specimen properties between reinforced and unreinforced waste materials. For example, the tensile strengths of reinforced and unreinforced polyamide 6.6 wastes were 84.2 and 165.2 MPa, respectively, which dropped to 38.0 and 97.0 MPa after the twentieth reprocessing cycle. Specimens from the reprocessing procedure have been investigated by Fourier transformed infrared spectroscopy in the spectral range of 400–4000 cm^{-1}. Due to mechanical stress between the rotating screw and plasticising cylinder, the reinforcements broke. The average length of the glass fibre was decreased as a function of the reusing cycle number from 1890 to 580 μm.

Keywords: polyamide, mechanical recycling, degradation, FTIR

Introduction

As a result of the continuously increasing trend in the demand for polymer worldwide, ever increasing amounts of waste polymers are generated. Polyamides (PA) are expensive engineering plastics, especially if they are reinforced. In particular, glass fibre composites are widespread among fibre-reinforced materials, as a results of their favourable mechanical and economical properties. The application of glass fibre results in improved tensile and flexural properties. Thermoplastic polyamides contain typically 10-30% glass fibres as reinforcements and are applied in many industries such as the automotive sector, civil engineering, packaging, and biomedical applications [1-9,11,16].

A grand challenge is the utilisation of plastic wastes. Polymer materials have a high resistance to different environmental and mechanical effects, which yields their favourable characteristics. Such advantageous properties can easily become disadvantage owing to waste polymers inability to decompose within a reasonable time frame. Unfortunately, a vast amount of polymer waste ends up to landfills, where non-degradable plastics together with other materials accumulate and pollute the environment [10,12-15]. The most studied recycling technologies are mechanical or chemical. Chemical recycling is a thermal process for cracking the long carbon chains of polymers into gases and liquid hydrocarbons in the absence of oxygen within in temperature range of 400-1000 °C. The other technology is mechanical recycling, when waste polymers are mixed with original granulates and newly shaped by injection moulding or extrusion. Mechanical recycling is a good option for reusing waste plastic, only when selectively collected polymers of high purity are available. Similarly to the chemical process, the polymer chains can degrade due to mechanical and thermal stresses during the pre-treating and shaping steps. Therefore, it is important to know the behaviour of polymer chains when exposed to thermal and mechanical stresses [9,11-16].

The goal of our work was to investigate the mechanical recycling of reinforced and unreinforced waste polyamides. Waste polymers have been reprocessed twenty times and the changes of mechanical amongst other properties were investigated. The effects of reinforcements on the loss of properties during the reshaping procedures were followed.

Experimental

Plastics

Glass fibre reinforced and unreinforced waste polyamides were used as raw materials. The main properties of the raw materials are shown in *Table 1*.

Table 1: The main properties of raw materials

	Reinforced	Unreinforced
MFI,* g (10 min)$^{-1}$	15.1	11.8
Ash content, %	30.6	0.03
Tensile strength, MPa	165.2	84.2
Charpy strength, kJ m^{-2}	13.4	12.1

* 275 ℃, 5 kg
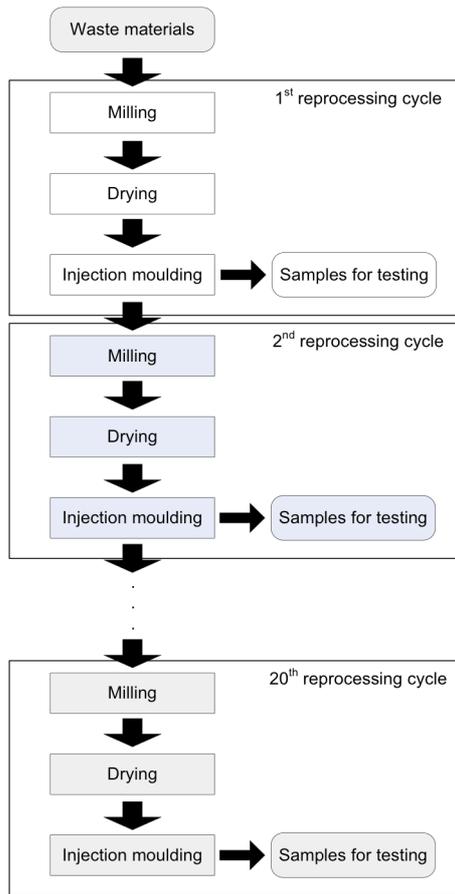
Figure 1: The main experimental steps

Figure 2: Tensile properties of reinforced and unreinforced polyamide 6.6 wastes: A) tensile strength, B) elongation and C) tensile modulus

Sample Preparation

Due to PA samples being sensitive to humidity, raw materials before processing and samples before testing were dried. Drying was carried out in an air circulated drying box at 90 ºC for 2 h. The injection moulding of the specimens was carried out by a microprocessor-controlled injection moulding machine. The dimensions of dog-boned injection moulded specimens were 3 mm × 10 mm in cross-section and 50 mm in length. *Fig.1* illustrates the main experimental steps.

Methods

The tensile and three point flexural properties (mainly stress and extension) (MSZ EN ISO 527-1-4:1999, MSZ EN ISO 14125:1999) were determined using an INSTRON 3345 universal tensile testing machine. The temperature and relative humidity in the laboratory were 23 ºC and 60% respectively, during the mechanical tests. Tensile tests were carried out at a crosshead speed of 80 mm min^{-1}. During the investigation of flexural properties, the crosshead testing speed was 20 mm min^{-1}. Samples were also investigated by infrared spectroscopy with a TENSOR 27 type FTIR spectrometer (resolution: 3 cm^{-1}, illumination: SiC Globar light, detector: RT-DLaTGS type) within the 400-4000 cm^{-1} wave number range. A CEAST Resil

Impactor was applied to measure the Charpy impact strength of the produced samples according to the MSZ EN ISO 179-2:2000 standard.

Results and Analysis

Tensile Properties

The tensile properties of reshaped specimens are summarised in *Figs.2A-2C*. The tensile strength of reinforced and unreinforced polyamide 6.6 wastes were 84.2 MPa and 165.2 MPa, respectively, which were reduced to 38.0 MPa and 97.0 MPa after the twentieth reprocessing cycle (*Fig.2A*). It is important to note that the decreasing trends were considerably different in the two cases: a fourth order trend line in the case of reinforced samples and a second order one in the case of unreinforced samples were found as shown in *Fig.2A*. A short section with a barely decreasing slope can be found both at the beginning (before the third cycle) and at the end (after the fifteenth cycle) of the trend line for

A

$$y = -0.0017x^4 + 0.086x^3 - 1.3572x^2 + 2.2191x + 247.71$$
$$R^2 = 0.9994$$

$$y = -0.0938x^2 - 1.6628x + 130.4$$
$$R^2 = 0.9988$$

△ Unreinforced ■ Reinforced

Flexural strength, MPa

Number of Cycles

B

$$y = 0.0426x + 4.875$$
$$R^2 = 0.9977$$

$$y = 0.0221x + 4.4028$$
$$R^2 = 0.9904$$

△ Unreinforced ■ Reinforced

Elongation, %

Number of Cycles

C

$$y = -0.037x^4 + 1.884x^3 - 29.702x^2 + 26.89x + 5599.2$$
$$R^2 = 0.9994$$

$$y = -0.0091x^4 + 0.4397x^3 - 8.058x^2 - 21.29x + 2632.2$$
$$R^2 = 0.9997$$

△ Unreinforced ■ Reinforced

Flexural modulus, MPa

Number of Cycles

Figure 3: Flexural properties of reprocessed reinforced and unreinforced polyamide 6.6 wastes A) flexural strength, B) elongation and C) flexural modulus

the reinforced specimens. In the case of unreinforced specimens a similar phenomenon was found only up to the ninth cycle.

The elongations of specimens are shown in *Fig.2B*. Owing to the rigid property of polyamides the relative elongations are rather small (below 3.8%). As results demonstrate the change in elongations was rather linear, but the gradient was different when the reinforced (tan α = 0.033) and unreinforced (tan α = 0.054) raw materials were investigated; because a slower ratio in change was found in the presence of glass fibres in polyamide wastes.

As demonstrated by the tensile strength, tensile moduli show a decreasing trend as well. Furthermore, the trend line is highly similar to the measured data for tensile strength. The E-modulus of reinforced and unreinforced wastes was 8219 MPa and 3226 MPa, respectively. Those were decreased to 3606 MPa and 1022 MPa according to the mathematical equations in *Fig.2C*.

△ Unreinforced
■ Reinforced

Charpy impact strength, kJ m^{-2}

Number of Cycles
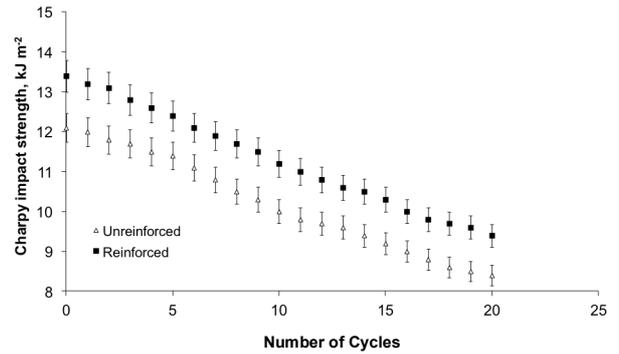
Figure 4: Charpy impact strength of reprocessed reinforced and unreinforced polyamide 6.6 wastes

Flexural Properties

The flexural properties of reshaped specimens are summarised in *Figs.3A-3C*. The change in both flexural strength and E-modulus followed a decreasing trend, while the trend in terms of elongation was rather increasing. The results were similar to the tensile properties discussed above. It is important to note that the shapes of trend lines were also similar to results demonstrated before. When reinforced polyamide 6.6 waste material was reprocessed the flexural strength, elongation and E-modulus changed from 248.5 MPa, 4.42% and 5622 MPa to 166.0 MPa, 4.86% and 3416 MPa, respectively. By taking into consideration the results of both tensile and flexural tests it can be concluded that the glass fibre reinforcement helps to keep the high values of strengths and E-moduli in the case of reinforced polymers at the start of property curves as a function of processing cycle number. Approximately after the third-to-fifth reprocessing cycles, the dropping ratio was increased step by step. Presumably this was the consequence of the glass fibres degrading into pieces. On the contrarily, a relatively constant dropping ratio was found in the unreinforced case.

Charpy Impact Strength

The Charpy impact strength test is one of the most widely specified mechanical tests of polymeric materials to provide information about impact properties. Results of Charpy tests were summarised in *Fig.4*. The change in Charpy impact strength followed a decreasing trend in both cases. Moreover, a similar trend in the changes could be found, because the gradients were 13.38 and 12.18 for reinforced samples and unreinforced specimens, respectively. The Charpy impact strengths were 12.1 kJ m^{-2} and 13.4 kJ m^{-2} without any reprocessing, and decreased to 8.4 kJ m^{-2} and 9.4 kJ m^{-2} respectively by the end of the procedure. Naturally the reinforced samples gave higher values of impact strength.
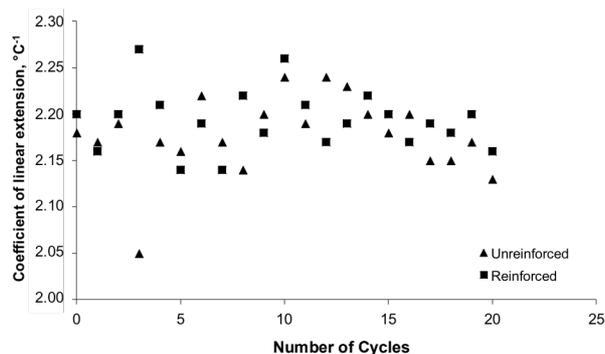
Figure 5: Coefficient of linear heat extension of reprocessed reinforced and unreinforced polyamide 6.6 wastes
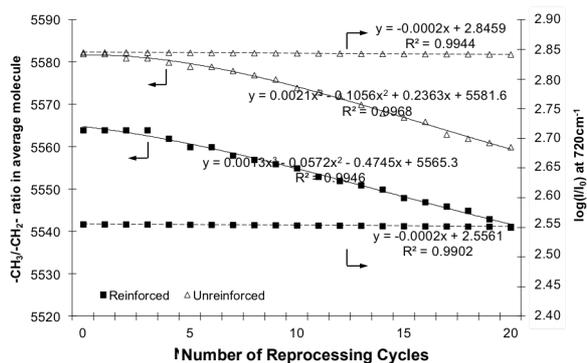
Figure 7: The change in the $-CH_3/-CH_2-$ ratio of an average molecule and the value of $\log(I/I_0)$ from FTIR spectra at 720 cm^{-1} as a function of the reprocessing cycles in the cases of reinforced and unreinforced polyamides

Figure 6: FTIR spectra of reinforced (A), and unreinforced polyamide (B)

Linear Extension Coefficient

The coefficient of thermal expansion describes the change in dimensions of samples against changing temperature at a constant pressure. Regarding polymers the linear extension coefficient is used to classify the materials. As it is known, thermal expansion decreases with increasing bond energy and the thermal extension of the amorphous phase is higher than that of the crystalline phase, because rearrangements in the structure can be achieved at the glass transition temperature. A change in ratio of amorphous and crystalline phases occurs. The coefficients of linear extension are demonstrated in *Fig.5*. Results demonstrate well that the coefficient changes randomly in both cases between 2.05 and 2.27 with approximately the same average value of 2.18 °C^{-1} for the reinforced and unreinforced specimens within 0.01 °C^{-1}. Presumably no rearrangements in the structures of the tested materials occurred during the reprocessing and the ratio of the amorphous to crystalline phases remained similar. In other words, the specimens a have relatively permanent stability of shape.

Fourier Transformed Infrared Spectroscopy

Samples from the reprocessing procedure were also investigated by Fourier transformed infrared spectroscopy (FTIR), which provides a measure of changes in molecular structures through molecular vibrational excitations. The FTIR spectra of reinforced and unreinforced waste polymers as a function of the reprocessing cycles are shown in *Figs.6A* and *6B*. By this method, modifications in the molecular structures of polymers could be observed as a result of changes in their chemical bonds. A typical spectrum of polyamide can be described as follows:

- stretching vibration of -CH$_2$- and -CH$_3$ groups between 2800–3000 cm^{-1}

- symmetric and asymmetric deformation stretching of -CH$_3$ groups between 1430 and 1470 cm^{-1} and 1365 and 1395 cm^{-1}

- $\beta_{asymmetric}CH_2$ of -CH$_2$- groups at 720 cm^{-1}

- bands at 3350, 1730 cm^{-1}, and 630 cm^{-1} as well referring to the N–H and C=O bonds from –NH– and –CO functional groups.

The methyl/methylene group ($-CH_3/-CH_2-$) ratio in an average molecule was calculated based on the FTIR peak intensities in the range of 2800–3000 cm^{-1}. Results are summarised in *Fig.7*. As shown, the ratio of methyl and methylene groups slowly decreased in both cases as a function of reprocessing cycles. This suggests that the numbers of methyl groups increased, while the methylene groups decreased in an average molecule. This is a consequence of polymer degradation, when the C-C bonds of main polymer chains are cleaved, which resulted in shorter chains containing less methylene groups.
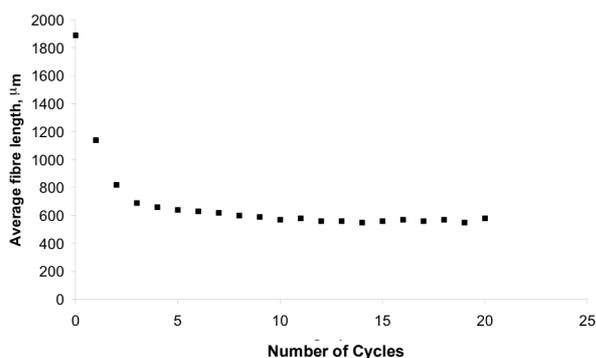
Figure 8: The average length of glass fibre reinforcements as a function of reprocessing cycles

Scanning Electron Microscope Graphs

As the polyamide matrix covered the fibres, it was difficult to determine their average length. Therefore to investigate the average length of the reinforcements and thus gain insights into their degree of degradation, the samples were burned and the remaining glass fibres were investigated by SEM techniques. The results are shown in *Fig.8*. As shown, the average length of the glass fibre reinforcements decreased as a function of the reusing cycle number. The length of glass fibres suddenly decreased from 1890 μm to 660 μm during the first four cycles of reprocessing. Then the average length exhibited a relatively constant value between 550 and 640 μm. The mean and standard deviation were 581 μm and 28 μm between the fifth and twentieth cycles, respectively. The cause for the degradation of the reinforcement fibres was the mechanical stress caused by the rotating screw and plasticising cylinder.

Conclusions

We investigated the reprocessing of waste glass fibre reinforced and unreinforced polyamide 6.6 materials. It was found that the mechanical properties of both wastes deteriorated as a function of reprocessing. For example, the tensile strengths of reinforced and unreinforced polyamide 6.6 wastes were 84.2 MPa and 165.2 MPa, respectively, which decreased to 38.0 MPa and 97.0 MPa after the twentieth reprocessing cycle. On the other hand, the trend lines of the changes were significantly different in the case of reinforced and unreinforced wastes. More gradually decreasing ratios could be measured regarding unreinforced plastics until the tenth cycle. Based on infrared spectra, it was concluded that the deterioration in mechanical properties presumably was the consequence of polymer degradation. The C-C bonds of the main polymer chains were broken and resulted in the formation of shorter polymer chains. The lengths of glass fibres suddenly decreased from 660 μm to 1890 μm during the first four cycles of reprocessing. Then, the average length exhibited a relatively constant value between 550 and 640 μm. The coefficients of

linear heat extension changed randomly in both cases between 2.05 and 2.27 °C^{-1}. The average value was 2.18 °C^{-1} for both reinforced and unreinforced samples within 0.01 °C^{-1}.

REFERENCES

[1] TROEV K., TODOROVA N., MITOVA V., VASSILEVA ST., GITSOV I.: Phosphorus-containing oligoamides obtained by a novel one-pot degradation of polyamide-6, Polymer Degrad. Stabil., *2006*, 91(4), 778-788

[2] PEDROSO A.G., MEI L.H.I., AGNELLI J.A.M., ROSA D.S.: The influence of the drying process time on the final properties of recycled glass fiber reinforced polyamide 6, Polymer Testing, *2002*, 21(2), 229-232

[3] GRÖNING M., ERIKSSON H., HAKKARAINEN M., ALBERTSSON A.C.: Phenolic prepreg waste as a functional filler with an antioxidant effect in polypropylene and polyamide-6, Polymer Degrad. Stabil., *2006*, 91(8), 1815-1823

[4] GOITISOLO I., EGUIAZÁBAL J.I., NAZÁBAL J.: Effects of reprocessing on the structure and properties of polyamide 6 nano-composites, Polymer Degrad. Stabil., *2008*, 93(10), 1747-1752

[5] BERNASCONI A., DAVOLI P., ARMANN C.: Fatigue strength of a clutch pedal made of reprocessed short glass fibre reinforced polyamide, Int. J. Fatigue, *2010*, 32(1), 100-107

[6] LA MANTIA F.P., SCAFFARO R.: Melt stabilization of wet polyamide 6, Polymer Degrad. Stabil., *2002*, 75(3), 473-477

[7] BERNASCONI A., DAVOLI P., ROSSIN D., ARMANNI C.: Effect of reprocessing on the fatigue strength of a fibreglass reinforced polyamide, Composites Part A: Appl. Sci. Manufact., *2007*, 38(3), 710-718

[8] HASSAN M.M., BADWAY N.A., GAMAL A.M., ELNAGGAR M.Y., HEGAZY E.S.A.: Studies on mechanical, thermal and morphological properties of irradiated recycled polyamide and waste rubber powder blends, Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, *2010*, 268(9), 1427-1434

[9] YU S., LIU M., MA M., QI M., LÜ Z., GAO C.: Impacts of membrane properties on reactive dye removal from dye/salt mixtures by asymmetric cellulose acetate and composite polyamide nanofiltration membranes, J. Membrane Sci., *2010*, 350(1-2), 83-91

[10] GHIDOSSI R., POUPOT C., THIBON C., PONS A., DARRIET P., RIQUIER L., DE REVEL G., PEUCHOT M.M.: The influence of packaging on wine conservation, Food Control, *2012*, 23(2), 302-311

[11] Hassan M.M., Badway N.A., Gamal A.M., Elnaggar M.Y., Hegazy E.S.A.: Effect of carbon black on the properties of irradiated recycled polyamide/rubber waste composites, Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, *2010*, 268 (16), 2527-2534

[12] Hassan M.M., Badway N.A., Gamal A.M., Elnaggar M.Y., Hegazy E.S.A.: Studies on mechanical, thermal and morphological properties of irradiated recycled polyamide and waste rubber powder blends, Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, *2010*, 268(9), 1427-1434

[13] Huang S.H., Hung W.S., Liaw D.J., Lo C.H., Chao W.C., Hu C.C., Li C.L., Lee K.R., Lai J.Y.: Interfacially polymerized thin-film composite polyamide membranes: Effects of annealing processes on pervaporative dehydration of aqueous alcohol solutions, Separ. Purific. Techn., *2010*, 72(1), 40-47

[14] Yu S., Liu M., Ma M., Qi M., Lü Z., Gao C.: Impacts of membrane properties on reactive dye removal from dye/salt mixtures by asymmetric cellulose acetate and composite polyamide nanofiltration membranes, J. Membr. Sci., *2010*, 350(1-2), 83-91

[15] Kiliaris P., Papaspyrides C.D., Pfaendner R.: Influence of accelerated aging on clay-reinforced polyamide 6, Original Research Article Polymer Degradation and Stability, *2009*, 94(3), 389-396

[16] Shen Z., Bateman S., Wu D.Y., McMahon P., Dell'Olio M., Gotama J.: The effects of carbon nanotubes on mechanical and thermal properties of woven glass fibre reinforced polyamide-6 nanocomposites, Composites Sci. Techn., *2009*, 69(2), 239-244