ABSTRACT
The paper introduces a decision support architecture with an integrated, self-building simulation module for the validation of the calculated manufacturing capacities, a-priori recognition of due date deviations and analysis of the effect of possible actions. In the underlying research special attention is devoted to the prediction and evaluation of the production on a daily, rolling time horizon basis (e.g., work in process (WIP) trajectories, machine utilizations).

The paper addresses the simulation module of a higher level integrated system in which the simulation model is automatically built on the basis of a real-time connection to Manufacturing Execution System. The main functionalities and advantages are highlighted through a real industrial case study.

Keywords: self-building simulation, production control, decision support, data mining

1. INTRODUCTION
In manufacturing systems, at the operational level, difficulties arise from unexpected tasks and events, non-linearities, and a multitude of interactions while attempting to control various activities in dynamic shop floors. The selection of the most appropriate production control decision for a given assignment, as well as the prediction of waiting times, workloads or utilizations of the resources are not trivial tasks, although they can be supported by simulation-based evaluations (Pfeiffer 2007, Bagchi et al. 2008, Watt 1998).

Therefore, based on previous results (Monostori et al. 2007), we propose a decision support architecture, in which the integrated, self-building simulation module can be applied for validation of the calculated manufacturing capacities, a-priori recognition of due date deviations and analysis of the effect of possible actions taken. In the research presented in this paper special emphasis is given to the prediction and evaluation of the production on a daily, rolling time horizon (e.g., Work in process (WIP) trajectories, machine utilizations).

The paper addresses the simulation module of the proposed architecture highlighting its main functionalities and advantages through a real industrial case study. An important issue regarding short-term (operational level) simulation is the automatic collection and definition of simulation input data. Therefore a new operation time definition method is presented in the paper, as well as a self-building simulation procedure is described in details.

2. PRODUCTION CONTROL AND SIMULATION – BACKGROUND

2.1. Simulation-supported production control
The discrete-event simulation (hereafter referred to as simulation) approach has been applied to decisions in scheduling and control, related to production applications (see e.g., Banks 1998, Law and Kelton 2000, O’Rielly and Lilegdon 1999). The simulation models that are used for making or evaluating these decisions (e.g., by projecting different key performance indicators, KPI-s) generally represent the flow of materials to and from processing machines and the operations of machines themselves (Rabelo et al. 2003). Potential problems can be identified and can be corrected using a simulation model. By far the most common use of simulation models is for operational decisions such as scheduling or dispatching (Law and Kelton 2000).

Simulation captures those relevant aspects of the Production Planning and Control (PPC) problem which cannot be represented in a deterministic, constraint-based optimization model. The most important issues in this respect are uncertain availability of resources, uncertain processing times, uncertain quality of raw materials, and insertion of conditional operations into the technological routings.

In simulation supported KPI evaluation, simulation is often used for evaluating the different scheduling or dispatching logics and methods. The usefulness of simulations lies in detecting and preventing the problems concerning KPI-s before they might occur at the shop floor. Thus, the key benefit of a simulation-based evaluation system is the feedback about system performance and system state (e.g., WIP, tool
utilisation) which, in turn, can be used for improving subsequent solutions.

2.2. Simulation-based schedule and control evaluation

As it is stated above, one stream of research in simulation-supported production control focuses on the simulation-based evaluation of schedules and scheduling rules. In these cases the main goal is to find solutions with the help of the simulation to daily scheduling issues including on-time order completion, priority changes, and unexpected changes in resource availability. Discrete Event Simulation (DES) helps a system engineer in detecting potential scheduling problems through checking the resource and schedule performances during the scheduling interval (shift, day, or week). The new alternative policies are then executed and performances of alternatives are compared. This process is repeated until a feasible and desired schedule is achieved. Indicated in another way, a schedule is created by simply simulating the execution of the factory and taking the recorded execution history as the schedule (Smith 1992).

Previous solutions introducing simulation-supported evaluation of schedules and scheduling policies are given in Cowling and Johansson (2002), Honkomp et al. (1999), Sabuncuoglu and Kizilisik (2003) and Vieira et al. (2000).

Watt (1998) presents a case-study where several information sources and applications are integrated. Simulation is applied in both off-line and scheduling modes. The current plant status and static data from the ManuMES are collected periodically and schedules are generated by a commercial scheduling package. Off-line simulations are performed in order to test what-if scenarios and reuse the same information for scheduling. New rules can be created and tested against history data. The improved rules are then applied in the scheduling system.

In several manufacturing areas the amount of products, product variants, resources, variation and fluctuation in the production processes does not facilitate predefined schedules to be adhered to. In these cases production is mainly controlled by production control rules (e.g., dispatching rules, for order management, and resource allocation, etc.) Thus, another direction of research does not focus on the schedule formulation but on the trajectories of the most relevant KPI-s during the simulation execution of the production. Solutions in the field of semiconductor manufacturing are presented in the followings.

Bagchi et al. (2008) describe a discrete event simulator developed for daily prediction of WIP position in an operational wafer fabrication factory to support tactical decision-making. The model parameters are automatically updated using statistical analyses performed on the historical event logs generated by the factory.


As it was shown in the previous literature review, related works describing some of the application areas, as well as the recent solutions of simulation in production control, simulation has been typically used for off-line decision making. Consequently, effective integration into the control process of production was restrained. One of the limitations of its use in on-line decision making is the considerable amount of time spent in gathering and analysing data. In quasi real-time control (hours, minutes), however, the three key issues are data acquisition, quick response and instantaneous feed-back. As a result, decision makers mostly apply simulation primarily for off-line decision support and not for the critical on-line decision making that may arise.

3. SELF-BUILDING SIMULATION SYSTEM

To support factory wide short-, midterm capacity as well as production control (utilization, WIP), scheduling and planning decisions, a simulation system has been developed – in form of a cooperative research project. This simulation system builds the discrete-event, object-oriented simulation model of the example shop-floor system automatically on the base of the data retrieved from the Manufacturing Execution System (MES) database.

This section presents the results of the systems’ analysis and the developed self-building simulation tool.

3.1. Novel simulation architecture – self-building simulation tool

At the beginning of the research activity related to this paper, in several preliminary examinations it turned out that the drawbacks of the existing simulation-based dispatching system are as follows:

- The input of the simulation model is oversimplified, collected and/or generated manually, which requires a huge amount of human effort;
- The data provision from shop-floor and engineering generates frequent errors;
- The data maintenance is poor;
- The responses in dynamic manufacturing environments are generally slow;

The main goal of development is the enhancement of the simulation-based analysis and dispatching system by eliminating the manual data collection through...
automatic interfaces, creating a more realistic model of the real factory and improving the dispatch logic inside the simulation model. Furthermore the self-building production simulation should provide both, prospective (e.g. locate anticipated disturbances, identify trends of designated performance measures), and retrospective (e.g. gathering statistics on resources) simulation functionalities. **Self-building simulation** means that the simulation model is build up by means of the combination of the MES data as well as the knowledge extracted from the MES data (e.g. resource and execution model). In addition to the automatic model building feature, main requirement of the solution is to minimize the response time of the experiments and to enable the quasi “real-time” applicability of the simulation.

### 3.1.1. Automated, component-based simulation model generation

In order to meet all the requirements and achieve the desired functionality for a flexible, self-building simulation system, a so-called **component-based simulation** method has been developed (Pfeiffer 2007). Resources, products, routings, production information, i.e., directly and indirectly usable data are gathered from MES database, and transformed as well as processed to the same form for all system components (e.g., for the simulation system, or for the Decision-maker to analyse KPI-s). Note that simulation relevant data (e.g. resource model, execution policies, process flow model) are redundantly stored locally in the simulation model.

![Figure 1: The structure and the data-flow of the new self-building simulation architecture (dotted lines represent information flow, while solid lines represent data flow)](image)

Figure 1 represents the dataflow of the new self-building simulation architecture. All the data necessary for simulation is retrieved from the MES database (represented on left side in Figure 1). The first interface splits the aggregated data of the MES into separated inputs and performs statistical analysis by applying previously developed algorithms. Production data are then further transferred and processed by the second interface into the simulation model (Prepare data in Figure 1). The simulation model keeps its own, internal, simulation-specific database to support fast response time. The exchange of the data is necessary only in the early state of the simulation when the automatic model building and data initialization is performed.

Data preparation is carried out before the overall simulation (production related data is refreshed on a weekly, while MES data for factory snap-shot is refreshed on a daily basis, see Figure 1). The redundant data storage in the simulation model is compensated by the advantage of the shorter response time. Modelling real production systems frequently brings up the problem of handling hundreds of resources in a simulation model. Having the modelling objects in hand, which were created on the base of the conceptual model, in our architecture the simulation model is created automatically based on the pre-processed data (phase b).

The automatic generation of the model is followed by initialization (phase c). There, besides classical parameter settings, the procedure involves the generation of input parameter specific model components (entities such as products, tools, machines...).
and the snapshot of the factory, detailed in Section 3.3). Contrary to the previous phase, this one is carried out for each replication. The simulation model incorporates a number of new dispatching rules with which the simulation runs can be manually initialized by simulation experts. In a following version of the simulation the automatic selection of proper dispatching rules and rule combinations will be implemented in order to achieve better performance in the shop-floor. The simulation is started on the base of these statistics by generating random production orders which cover the product type distribution calculated from the MES database. Naturally, instead of randomly generated orders, the users of the simulation can also provide the input for the simulation model on the base of real customer order data.

The simulation runs are repeated until the required number of replications is obtained (phase d). Each replication is a terminating, non-transient simulation run.

In the last phase, the results are evaluated (critical values for defined KPI-s) and the results of the evaluation process are interpreted by the Decision-maker (e.g., planner or dispatcher) who is responsible for taking the necessary actions. Several simulation results and statistics are calculated inside the simulation model and a graphical user interface (GUI) is provided for the visualization via a web browser of both, input settings, and statistical results. The first version of this self-building simulation system is developed in Plant Simulation v8.1 software, while the GUI is implemented in a web application.

3.1.2. Production information – data in the MES database
The production environment to be modelled by the simulation covers a whole production section of a real industrial factory. The factory produces several different products which are identified by different product types. Each product type is assigned to a routing. The sequence number is the position of a defined operation in the routing. In order to have a better traceability of products moving through the factory a number of operations are logically grouped into so called groups. Regarding resources in the factory we can differentiate between operators and machines. Products are processed on machines with predefined tools and machines are grouped into cells.

Operators are responsible for loading, starting, finishing, as well as unloading the machines. The data in the MES database are collected manually by the operators of the machines during the manufacturing process in the shop-floor. The products are transferred from station to station in lots. When a product arrives to a station the operator of the station registers the arrival and the product is waiting for the manufacturing in the queue of the station. Depending on the type of the tool different products might be processed together.

When an operation is completed the operator of the station registers the completion of the work. As a consequence the MES database contains a large list of data records about the life-cycle of the products cataloguing the entrance and completion events.

The data about the resources, their types and availability, routings, process times are also calculated from the MES database. MES-based statistics are collected about the most important product types and their distribution in the production order.

3.2. Defining operation times on MES data
In simulation systems representing a complex, wide scale production system, exact processing times are crucial for successful and credible simulation results.

Defining process times for simulation models based on logged production (e.g., MES) data is a well known and widely used method. Bagchi et al (2008) present a linear regression method for calculating process times based on raw process times (RPT) collected for single, batch and sampling tool types. Sivakumar and Chong (2001) present a case study, where the theoretical process times are defined by the mean values, however the authors state that based on the wide distribution in theoretical process times, theoretical ratio based on mean is indefinite.

![Figure 2: Frequencies of raw process times for a defined operation subtype on batch tool machines in seconds](image)

In Figure 2 an example is given, representing the characteristic of the raw process time distribution. The main goal is to define the so called basic process time for the given operation and tool relations. Since several factor influences the raw process time as for instance different waiting times in the input buffer of tools, different operators, etc., the most relevant way is to scrape the raw process times from the effect of disturbing factors, and thus, to identify the relevant lower bound (excepting unnecessary, problematic data), i.e., the shortest possible raw process time of the data presented in Figure 2. It is clear that by calculating the mean (31193 s) or the mode (29520 s) values for the data set does not provide the necessary lower bound.

Thus, as one of the most important issues regarding the self-building simulation system, a significantly more effective method (compared to e.g., mean or modus statistical methods) had to be developed in order to have reliable and up-to-date process times in the simulation. It is also important that the calculated process times will be applied in the near future for static capacity calculations in the factory.
Intensive research is still ongoing in our laboratory to calculate the raw process times. The system presented in this paper applies the preliminary results of the first algorithms. The detailed algorithm will be published soon in an upcoming paper of the authors.

3.3. Simulation model - Creating snapshot of the production

In order to have as much precise simulation results of a daily WIP prediction as possible, the initial state of the simulations (from which the system starts evolving) is critical. Therefore, a snapshot of the production (WIP, status of resources, etc.) is necessary (Figure 3).

![Proposed architecture of the simulated execution of a MES section, the formulation of the WIP snapshot and the prospective simulation](image)

Figure 3: Proposed architecture of the simulated execution of a MES section, the formulation of the WIP snapshot and the prospective simulation

In Figure 3 the blue box represents the operations executed during the Log Simulation, while green boxes represent the steps to be processed in the routing. The main challenge is the transition phase at time point $T_0$, where the execution of the log-based operations should be changed to routing-based operations for each unit to be processed.

We propose the following simulation procedure to ensure a short warm-up phase and a reasonable initial state of the production status.

- The user should define the simulation time horizon
  - Starting time ($T_0$) at which the prediction starts
  - Length of the WIP prediction phase (e.g. 1 week), i.e., $T_0 - T_{\text{f}}$
- Estimation is given on the length of the play-back period, required for the snapshot formulation ($T_0 - T_1$), in order to have the actual WIP state of the factory (warm-up phase)
  - Define the level of significance and the number of simulation runs required
- Play-back of the log (Log Simulation)
- From the snapshot point in time (T0) execute the WIP prediction phase (several independent simulation runs, Prospective Simulation)
- Evaluation of the results, notify user

In the simulation model we do not simulate the movement of individual products, but the movement and processes done on the lots. Therefore, the selected section (defined by $T_0 - T_1$) of the log (for creating the log file in the simulation) is decomposed and forwarded to lot objects in an aggregated way. This procedure improves the speed of the simulation, while this lower level of model granularity does not significantly influences the output quality.

4. PRELIMINARY SIMULATION RESULTS

The main goal of the simulation studies are the prospective analysis of the WIP movements in the factory, as well as the monitoring and reporting on bottle-neck machines and machine groups (for capacity analysis).

Our solution is distinct in the sense that the extent of modelling is relatively large (whole factory) while the data available for model formulation is limited only to the – continuously updated and statistically analysed – MES data (no conventional process and resource data is available).

The production system to be modelled has the following main features:

- several million rows in the MES database, 20-50,000 new rows every day;
- 3000-4000 products for snapshot (Factory WIP);
- approx. 500 resources;
- 500-1200 operations in a routing for a defined product type;
- avg. lead time 20-40 days;
- ~10-15% rework process.

The section presents an example of the validation process of the simulation model (Section 4.1), and preliminary results of the simulation experiments. Section 4.2 introduces the steady state analysis of system performance, while in Section 4.3, results on testing improved execution and sorting logics are described.

4.1. Validation – comparison of throughput

In order to have credible results computed by the simulation model, a comprehensive validation procedure is mandatory. In the followings a short example is given about the validation of throughput values.

One stream of the validation procedure of the proposed system is based on the comparison of the simulated WIP prediction results (simulated log) with the real original data (Figure 4). This validation procedure serves as a feed-back for the iterative fine-tuning (trace technique, Law and Kelton (2000)) of the execution policies, tool models, process times and dispatching rules applied in the simulation.
In Figure 4 the results of a comparison for one week period is highlighted. If the total number of products in the system (WIP) is considered as 100%, then the ratio of exiting products all together in the simulation for the one week time horizon in consideration is 1.222%, while the real, log based ratio is 1.306%. The difference between the predicted and the real total throughput is relatively low (6.38%), however the distinct steps in the log based curve can not be represented exactly with the simulation.

4.2. Testing production control rules - steady state performance

Regarding the shop-floor level control of the production, several control rules are implemented and used in the factory. One of the most relevant groups of these control logics is the work-load balancing dispatching rules of the machines. These rules define the logic of choosing the successor machine in case of the next operation can be processed on more than one (alternative) machine. At the current phase of implementation the following rules are available for the simulation:

- **Random**: the simplest logic to send products randomly to the next tool.
- **Min queue**: this logic prefers the machine where the queue is the minimum.
- **Relative occupation (Tool)**: this logic gives priority for the machine that has the Tool with the lowest relative occupation.
- **Relative occupation (Buffer)**: similar the one above, but considers the Buffer.
- **Next free (Tool)**: this logic prefers the machine where the Tool is free. It chooses randomly if there is no free Tool among the alternative machines.
- **Next free (Buffer)**: similar the one above, but considers the Buffer.
- **Min wait time**: in this logic a method calculates the time remaining to the start time of the process for the product on every possible machine. The product is sent to the Tool where this time is the minimum.

In order to be able to analyse the effect of dispatching rules on bottle-neck machines, first the long-term, steady-state analysis of the simulation model was performed (Table 1). It can be stated that the different dispatching rules significantly affect the systems’ throughput performance, e.g. the rule *Min wait time*.

Table 1: The average number of finished products (simulation time horizon is 50 days, number of replications is 100)

<table>
<thead>
<tr>
<th>Dispatching rule</th>
<th>No. of finished products</th>
</tr>
</thead>
<tbody>
<tr>
<td>Next free (Tool)</td>
<td>2619</td>
</tr>
<tr>
<td>Relative occupation (Tool)</td>
<td>2188</td>
</tr>
<tr>
<td>Relative occupation (Buffer)</td>
<td>2434</td>
</tr>
<tr>
<td>Min wait time</td>
<td>2865</td>
</tr>
</tbody>
</table>

As an important output of the self-building simulation system, situation-dependent selection of the most relevant control rules are intended to be tested and analysed. Therefore, contrary to the above described analysis, not only the steady-state performance of the defined control rules is of interest, but the short-term effect regarding the WIP in the input buffers of the machines. Moreover, the dynamic behaviour, e.g., the movement of these WIP “waves”, from machine to machine has to be considered.

In Figure 5 and Figure 6 the predicted trajectories of the WIP are presented for the selected bottle-neck machines, simulated for the dispatching rules *Min wait time* and *Random*, respectively. Based on the simulation results it can be stated that, the control rules with a reasonable high performance for steady-state studies, not always perform similarly good regarding short-term decisions (e.g., for eliminating WIP in buffers).
4.3. Suggestions for improving utilisation of parallel machines

During the analysis of the real systems’ data in the MES database, it turned out that the actual load of some machines, and thus the utilization is constantly low.

These machines are of principally one type, namely, special parallel machines, characterised by a certain maximal load capacity and with high operation costs. In case of the processing of the products starts, no more products can be loaded into the machine unless the processing is finished. Thus, the main goal is to minimize the idle time of these machines and to maximize the utilization (load). Therefore, three different loading logics have been developed and compared through simulation experiments. The three logics are as follows:

- **FIFO (or no group)** means the simplest logic: product lots are processed based on the FIFO rule;
- **Group** logic regroups the product lots, based on the operation type (OpType) required, and then it sorts according to the number of products in the new group. The order of processing is FIFO;
- **Group and sort** logic first regroups the product lots as described in Group. Each group (based on the OpType) has a predefined “starting level”. It defines a loading level at which processing is allowed to be started. A **waiting time** is also assigned to each OpType. The group which first exceeds the starting level or exceeds the waiting time will be first processed on the machine.

The upper chart in Figure 7 presents the case, where No group (FIFO) rule is applied to load the parallel machines, while the lower chart presents the case, where the special sorting logic (Group and sort) is applied in order to have a higher machine utilization ratio. In the first case, the machine does not wait for OpType1 operations to load the machine up to full capacity, but starts processing immediately after receiving products for OpType1, or OpType2. Contrary, in the second case, the processing of products requiring OpType1 are postponed, because the minimum level assigned to this type of operation is not exceeded.

The results of simulation experiments regarding the effectiveness to the new sorting logics are presented in Figure 8 and Figure 9 for two selected different machines. It is clear that the new Sort and group algorithm has a positive effect on the WIP level of machine input buffers.
5. SUMMARY AND FURTHER STEPS
The features provided by the new generation of simulation software facilitate the integration of the simulation models with the production scheduling and control systems. Additionally, if the simulation system is combined with the production database of the factory, it is possible to instantly update the parameters in the model and use the simulation parallel to the real manufacturing system supporting and/or reinforcing the decisions on the shop-floor.

However, some tasks are still under development stage (e.g. calculation of reliable process times), in this paper the authors presented a novel solution to build simulation models from the MES database. Moreover, the initial validations, which were carried out with real industrial data, showed that the behaviour and the results of the simulation model are close to the ones of the real system.

As a further step, regarding the usage of our self-building simulation approach, we propose the re-initialization and memorization of sequential simulation runs in a rolling time intervals. In case of high difference arises between the simulated and the real production KPI-s (WIP, waiting times, etc.), an examination is initiated in the log of the last interval to identify the root cause of the deviation.

The main challenges related to this rolling time horizon analysis are the identification of relevant rolling interval (i.e., the frequency of simulation experiments), defining significant ΔKPI-s and the definition of look-ahead horizon.

With predefined simulation protocols, this method will result in an easy-to-use decision support tool for shop floor engineers, which gives prediction of future events at a time by which necessary actions can be taken in advance.

ACKNOWLEDGMENTS
This research has been supported by the Hungarian Scientific Research Fund (OTKA) grants “New mathematical models and methods for integrated production planning and scheduling” K-76810, and “Production Structures as Complex Adaptive Systems” T-73376.

REFERENCES


Watt, D.G., 1998. Integrating simulation-based scheduling with MES in a semi conductor fab, in...
AUTHORS BIOGRAPHY

András Pfeiffer, earned his PhD in 2008 at the Budapest University of Technology and Economics. Currently he is a senior research fellow at the Engineering and Management Intelligence Laboratory of the Computer and Automation Research Institute, Hungarian Academy of Sciences (SZTAKI). His current interest includes decision support in production planning and control, as well as the simulation and emulation modelling of complex production systems, self-building simulation systems.

Botond Kádár is a senior researcher at the Engineering and Management Intelligence Laboratory of the Computer and Automation Institute of the Hungarian Academy of Sciences (SZTAKI). He obtained his MSc and Ph.D. degrees at the Budapest University of Technology and Economics, Hungary, in 1993 and 2002, respectively. His current interest includes production control, simulation and multiagent approaches for production engineering and manufacturing systems and he is involved in several research and development projects from these fields. Dr. Botond Kádár is author or co-author of 70 publications with over 120 citations.

Marcell Szathmári, graduated in 2006 at the Budapest University of Technology and Economics. Currently he is a research associate at the Engineering and Management Intelligence Laboratory of the Computer and Automation Research Institute, Hungarian Academy of Sciences (SZTAKI). His current interest includes tracking information systems and automatic identification technologies.

Gergely Popovics, graduated in 2008 at the Budapest University of Technology and Economics. Currently he is a research associate at the Engineering and Management Intelligence Laboratory of the Computer and Automation Research Institute, Hungarian Academy of Sciences (SZTAKI). His current interest includes the simulation modelling of complex production systems and automatic identification technologies.

Zoltán Vén, graduated in 2006 at the Budapest University of Technology and Economics. Currently he is a research associate at the Engineering and Management Intelligence Laboratory of the Computer and Automation Research Institute, Hungarian Academy of Sciences (SZTAKI). His current interest includes reconfigurable manufacturing systems, as well as the simulation and emulation modelling of complex production systems, self-building simulation systems.

Prof. László Monostori has been with the Computer and Automation Institute of the Hungarian Academy of Sciences (SZTAKI) since 1977, now he serves as Deputy Director Research and Head of the Engineering and Management Intelligence Laboratory. He is also full time professor at the Department of Manufacturing Science and Technology, Budapest University of Technology and Economics. He is a Fellow and Council Member of the International Academy for Production Engineering (CIRP); Full Member of the European Academy of Industrial Management (AIM); Chairman of the Coordinating (CC) on Manufacturing and Logistics Systems, International Federation of Automatic Control (IFAC), and Chairman of the TC on Technical Diagnostics, International Measurement Confederation (IMEKO). He is Chairman of the Editorial Committee of CIRP Annals; Editor-in-Chief of the CIRP Journal of Manufacturing Science and Technology; Associate Editor of Computers in Industry, as well as the Measurement, and member of the editorial boards of other international scientific periodicals. For his research achievements published in more than 340 publications resulted in about 1500 independent citations and for his development activities – among others – the Dennis Gabor Prize was given to him in 2004. Prof. Monostori is a member of the Hungarian Academy of Engineering.