Proceedings of the 2003 Winter Simulation Conference
S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, eds.

AN APPROACH TO ROBUST LAYOUT PLANNING OF AMHS

Roland Sturm
Joachim Seidelmann
Johann Dorner
Kevin Reddig

Fraunhofer IPA
Dept. Cleanroom Manufacturing
Stuttgart, GERMANY

ABSTRACT

The simulation-based layout planning of automated material handling systems (AMHS) for microelectronics and semiconductor manufacturing demands adequate simulation models. An approach for measuring and quantifying the AMHS layout performance of alternative planning variants is required. Fraunhofer IPA has developed simulation methods and a three level approach for calculation of AMHS performance metrics. This approach is very efficient when comparing alternative planning variants, although the difference in the configuration change is very small. The paper outlines the planning approach for two typical AMHS designs used for interbay transportation in 200mm wafer fabs. The models used are generic and can be adapted easily to different AMHS solutions.

1 INTRODUCTION

Planning, dimensioning and configuration of automated material handling systems (AMHS) in semiconductor manufacturing for robust operation is a sophisticated task and requires the interdisciplinary work of IC manufacturers and AMHS suppliers. Due to rapid changing process flows, product mix and adjustment of equipment capacity, installed AMHS often require reconfiguration, expansion and layout changes. The focus of this paper is to outline an evaluation approach when planning a new AMHS installation or modification of an existing AMHS with regard to good performance and robustness of the system. The challenge of AMHS planning is increased for flexible AMHS interbay layouts where each stocker to stocker connection can be reached by alternative travel paths. For 300mm wafer fabs with automated tool to tool transportation scenarios this challenge is likewise seen for intrabay and interbay AMHS planning.

We present some simple metrics which help to identify weak points in AMHS layouts which can be investigated based on output data from discrete event simulation or logged events from the real system. The metrics can also be applied for the daily or weekly performance monitoring of an AMHS installation.

In most wafer fabs the position of process equipment and stockers are determined by the fab layout. The Manufacturing Control System (MCS) which is the control software of an AMHS installation, is designed for a specific transportation system and can only partly be adapted to individual needs of a wafer fab. Therefore, the common way to adapt the AMHS to changes in the process flow or process equipment capacity is by altering AMHS components, such as track layout, turntables, number of vehicles or configuration of parking loops and adaptation of stocker shelf capacities.

Discrete-event simulation helps to understand the robustness of an AMHS design variant in terms of delivery times, utilization of AMHS components (critical stockers and vehicle quantities) and zones of potential vehicle congestion.

2 PROBLEM STATEMENT

The evaluation of alternative AMHS designs is often done based on global performance metrics such as delivery time or throughput averages over a long observation period. Consequently, local problems such as temporary vehicle congestion or stocker overflow is difficult to evaluate based on simulation output statistics such as global average metrics. Congestion measures have been extensively used for the evaluation of roadway systems (Beamon 1999). However, the application of congestion measures to material handling systems, especially during the planning phase of an AMHS, has been limited. This paper addresses (1) an easy to implement approach of measuring AMHS congestion in a wafer fab based on simulation results and (2) investigates a methodology for robust AMHS configuration planning based on the presented metrics.

3 SIMULATION MODEL

Our planning approach is discussed based on simulations performed with a generic AMHS simulation model (AutoMod V10.0) implemented by Fraunhofer IPA. The path mover system was used to model an overhead monorail system (following the AeroTrak System of Brooks Automation) which connects 7 stockers located in 6 process areas (Figure 1). The model can be adapted easily to systems of other AMHS suppliers. The vehicle routing logic was customized in the AutoMod model as described by Brooks Automation (Norman 2002).

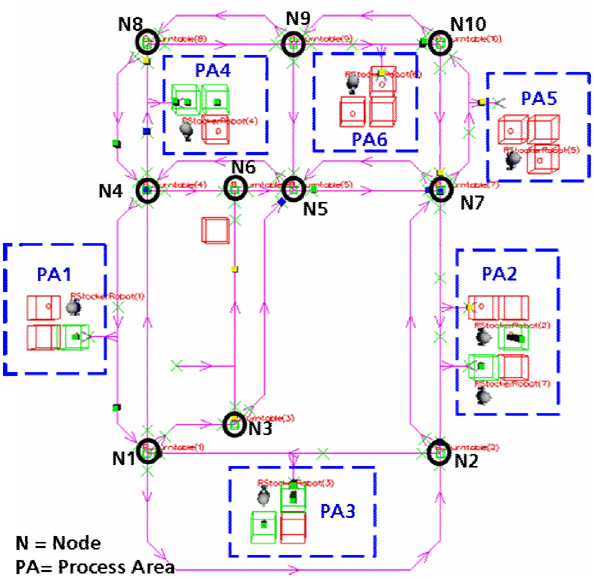

Figure 1: AMHS Simulation Model

The lots in the model are driven by a reentrant process flow, the lot delay time in the process areas and the batch size if the process step requires several lots to be processed together. At the beginning of each shift (8 hrs) a certain amount of wafers (wafer start) combined in lots with the standard 25 wafers each are released into the first process area. The start rate drives the throughput of the system. Each stocker has a limited shelf capacity and an alternate stocker. Once the stocker shelf utilization comes to 85%, lots are transported to the alternate stocker.

We customized the simulator so that log files are written during the simulation run. For instance, a node crossing log file reports the timestamp each time a loaded or unloaded vehicle crosses a control point along with the status of the vehicle (retrieving, delivering or going to park) as depicted in Figure 2. Additionally, a move log file is written which reports arrival times of lots during the transportation and storage sequence of lots in the AMHS (Table 1).

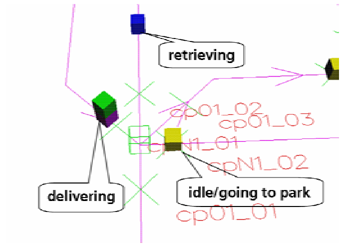

Figure 2: Turntable and Vehicles of the Overhead Monorail System in the Simulation Model

Table 1: Timestamps of Move Log File

Time-stamp	EVENT
TS1	Operator stores lot at IO-Port (source stocker)
TS2	Lot is on shelf in source stocker/transport job requested
TS3	Vehicle is allocated to pick up the lot
TS4	Allocated vehicle arrives at stocker monorail port – transportation to destination stocker starts
TS5	Arrival of the loaded vehicle at the monorail port of the destination stocker
TS6	Lot is on shelf in the destination stocker
TS7	Lot is on IO-Port of destination stocker (operator retrieves lot for processing)

4 AMHS PERFORMANCE METRICS

4.1 Traditional AMHS Performance Metrics

The performance of an AMHS installation or an AMHS simulation study is usually evaluated by the following metrics:

- Throughput
  - Moves per hour (MPH)
- Delivery performance
  - Average delivery time (DT) between stocker connections (TS6 - TS2)
  - Average transportation time (TT) between stocker connections (TS5 - TS4)
- System utilization
  - Vehicle utilization
  - Stocker robot utilization
  - Stocker shelf utilization
  - Turntable utilization.

MPH, DT and TT are usually calculated as global average of the system in AMHS simulation models. In a vehicle based system we define the transportation time TT as the time from the pick-up of a load at the source stocker to the arrival time of the loaded vehicle at the destination stocker. The delivery time is defined as the time from the first move request (vehicle is ordered for a retrieve job at the source stocker) to the arrival time of the load on the shelf of the destination stocker. System utilization metrics are calculated for each AMHS components.

When comparing AMHS layout variants or AMHS configurations, it shows that differences of global measures, for instance TT, are often negligible. Therefore, the comparison of AMHS layouts, especially when comparing large systems is very difficult based on global measures. Local problems of an AMHS design are often derived from utilization metrics, for instance robot utilization or turntable utilization measured in vehicle crossings.

Another possibility is the calculation of MPH, DT and TT of each stocker to stocker connection or even each node to node connection. A node can be a turntable or a stocker monorail port (PPort). However, the high number of possible segments requires comprehensive spreadsheet calculations or database operations.

4.2 Congestion Metrics

Beamon (Beamon 1999) describes a congestion index which can be used to describe the overall congestion level of a material handling system. We adapted this measure to the evaluation of AMHS congestion in a wafer fab:

$$C_{TTindex}(src,dest) = \frac{TT(src,dest)}{TT_{\min}(src,dest)} \quad . \quad (1)$$

$C_{TTindex}$ is the congestion index of all transports performed from a source node *(src)* to a destination node *(dest)* over a certain observation period. *TT(src,dest)* is the average of all observations from the source to the destination node, whereby $TT_{min}(src,dest)$ is the shortest travel time observed during the observation period. The congestion index has a range of [1.0,∞). In other publications TT_{min} is also referred to a technical transportation time, if there was no congestion at all. The congestion index can be calculated for stocker to stocker connections or for turntable-to-turntable along the travel paths in the layout. The latter helps to identify critical path segments in the AMHS layout and turntables with high queue times of vehicles. For node-to-node congestion index we only consider vehicles in the “delivering” or in the “retrieving” state. The congestion index of stocker-to-stocker delivering times is calculated in an analogous manner.

For the evaluation of the congestion index, it is also important to know how many moves are observed in each specific node-to-node connection. For instance, a high congestion index in a node-to-node connection which is based on few moves only has not the level of significance compared to another connection with very high move rates.

Another metric, which combines in a specific node-to-node connection the transportation time and the throughput view is the TT/Moves ratio:

$$\tan\alpha(src,dest) = \frac{TT(src,dest) - TT_{\min}(src,dest)}{Moves(src,dest) - 1} \quad . \quad (2)$$

The metric tanα represents the gradient of a node-to-node (source to destination) connection. High gradients indicate very sensitive connections and potential bottlenecks of the system. Temporary surges of the move rate may result in a very high increase of the transportation time when the connection indicates a high gradient. Figure 3 shows this measure represented by a scatter plot of TT over number of observed moves. The TT/Moves gradient of an AMHS design 1 (Var1) and design 2 (Var2) can be compared directly.

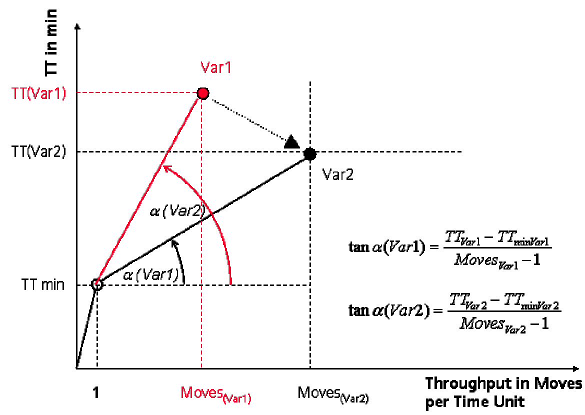

Figure 3: Comparison of a Node-to-Node Connection of two Layout Variants with the TT/Moves Ratio

4.3 Application of Metrics for Robust Planning

We follow a three level approach for the evaluation of AMHS configurations, for instance alternate layouts:

4.3.1 Global Move Performance

First we calculate global metrics for DT, TT and MPH without any resolution from which source to which destination a move took place based on our move log file. For example, the global average delivery time is calculated with

$$DT = \frac{\sum_{j=1}^{m}\sum_{i=1}^{n} DT(i,j) * Moves(i,j)}{\sum_{j=1}^{m}\sum_{i=1}^{n} Moves(i,j)} \qquad (3)$$

where *n = number of source stockers* and *m = number of destination stockers* and *Moves (i,j)* is the number of transported loads from source *i* to destination *j*.

4.3.2 Local Move Performance

Here we calculate DT, TT and MPH for all source stocker to destination stocker connections. Additionally, we calculate these metrics from destination point of view. Destination metrics represent the view of the operator at the destination stocker waiting for lots to be processed in the process area. For example the destination delivery time for each stocker can be determined with

$$DT_{dest}(j) = \frac{\sum_{i=1}^{n} DT(i) * Moves(i)}{\sum_{j=1}^{n} Moves(i)} \qquad (4)$$

where j is the destination stocker and *n=number of all source stockers to destination stocker j.*

4.3.3 Congestion Analysis

Finally we calculate node-to-node congestion metrics based on node crossing log files as described in formula (1) and (2).

With the three level approach alternative AMHS designs can be compared efficiently. Furthermore, it helps to identify the AMHS configuration with the best performance for the investigated manufacturing environment.

5 SIMULATION-BASED AMHS LAYOUT COMPARISON

We describe the described three level evaluation approach exemplary for the comparison of the AMHS layout shown in Figure 4 and Figure 5. These AMHS layouts are not from a real wafer fab. However, AMHS layouts with a much higher complexity were investigated with the same approach as presented here within Fraunhofer research projects. Results from these projects cannot be presented here due to non-disclosure agreement with our project partners.

Simulation experiments with Layout 1 showed temporary queuing at the PPorts of stocker 2 (S2) and S7 in the animation view of the simulator. S2 and S7 are parallel stockers of process area 2 connected to the same path segment. Loaded vehicles which have to travel to S7 are often blocked by a queue of vehicles at the PPort of S2. However, there is a “round the block” function. This means that vehicles waiting for delivery or pick-up at S2 are bumped to N2 after a pre-defined waiting time (here: 20 seconds), dispatched to N7 and finally queue behind the waiting vehicles at S2 again. Consequently, we wanted to investigate if a separation of S2 and S7 in segregated loops of the monorail track is beneficial for the AMHS performance and robustness. The layout change from Layout 1 to Layout 2 requires more monorail track only and no additional turntable is needed.

Both layouts were simulated with a simulation time of 1000 hours. The log files from the simulation runs were transferred to a data base and all calculations were performed with database queries.

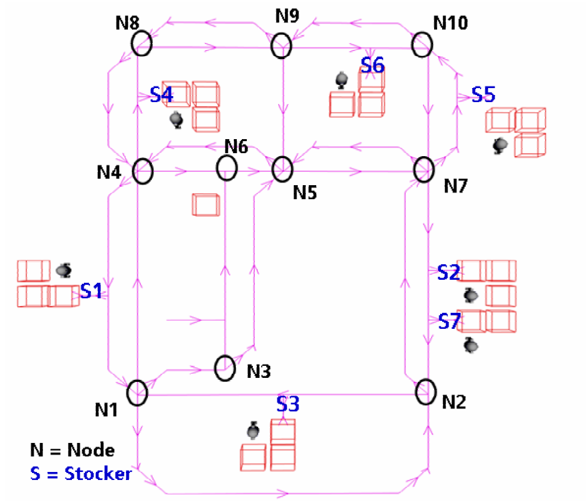

Figure 4: AMHS Layout 1

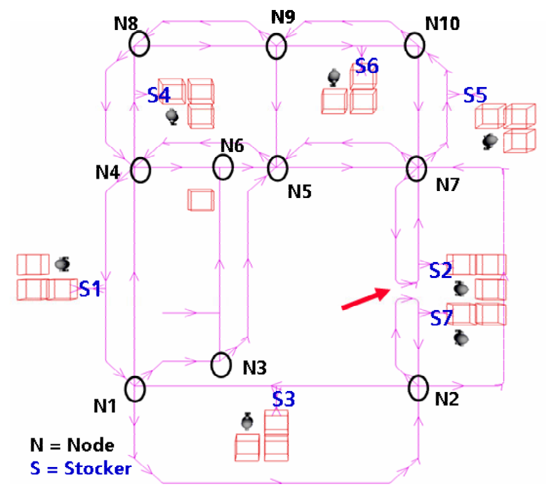

Figure 5: AMHS Layout 2 with Separated PPorts S2 and S7

First we calculated from the move log file global measures for the average delivery time, average transport time and MPH of the system. Table 2 shows the results. DT for Layout 2 is increased significantly. The utilization metrics of stocker robots, vehicles show no major differences between the two layouts (not shown here). The longer delivery times may result from empty vehicles which have to travel a longer distance from the parking lot, which is located on the path section between N3 and N6, to pick up loads at S7. However, this is just a first assumption which cannot be proved based on the global metrics. The global congestion metric $C_{TTindex}$ (node-to-node), which is the average over all node-to-node connections shows a smaller value for Layout 1 compared to Layout 2 (Table 3). This indicates, that a higher congestion more traffic is observed over the total period in Layout 2.

Table 2: Global move performance metrics

LAYOUT	MPH	DT	TT	MinDT	MinTT
Layout 1	89.78	39.44	4.55	2.95	1.7
Layout 2	89.47	52.45	4.94	2.95	1.7

Table 3: Global congestion metrics

LAYOUT	$C_{DTindex}$ (Stocker-to-Stocker)	$C_{TTindex}$ (Stocker-to-Stocker)	$C_{TTindex}$ (Node-to-Node)	TT/MPH (Node-to-Node)
Layout 1	13.37	2.68	6.24	0.00009
Layout 2	17.78	2.91	6.54	0.00009

On the next level we compare delivery performance and throughput metrics for all process area to process area connections separately. This can also be done on a stocker-to-stocker basis, which is for our layouts nearly the same because we only have process area 2 which contains S2 and S7.

In the scatter plot shown in Figure 6 the metrics DT(src,dest) and the MPH(src,dest) for layouts 1 and Layout 2 are compared. Layout 2 shows increased DT for all active process area to process area connections. Especially, much higher DT are observed for the process area connections 3→4 and 2→3 shows a. Throughput values of the direct comparison of Layout 1 and Layout 2 stay the same.

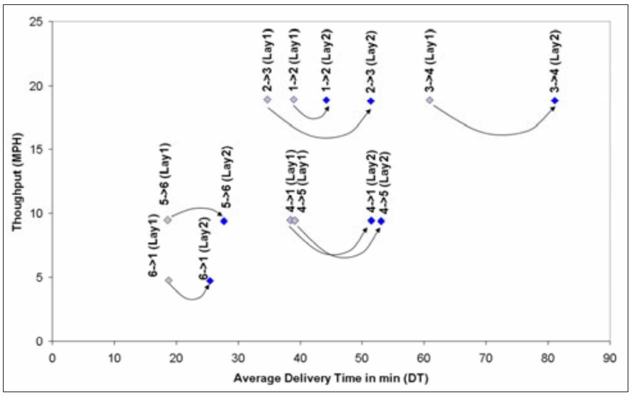

Figure 6: Process Area to Process Area Delivery Time DT(src,dest) - Layout 1 vs. Layout 2

The grouping of the delivery times by process area from a destination point of view is depicted in Figure 7. This shows also that the load transfer to all process areas indicates a longer delivery time for Layout 2. Particularly, $DT_{dest}(4)$ shows an increased value. This is surprising, because process area 4 with S4 is nearly at the opposite corner of the layout and is influenced to a high extend by the PPort separation of S2 and S7 in Layout 2.

Figure 8 shows the transport time metrics on process area level. A smaller TT(6,1) and TT(4,1) show smaller values and perform better in Layout 2. However, TT(2,3) is increased from 5 to 8 minutes in average. This result was expected, because in Layout 2 a vehicle has to travel a much longer path from the source S2 to the destination S3 (S2, N7, N5, N6, N4, N1, N2, S3).

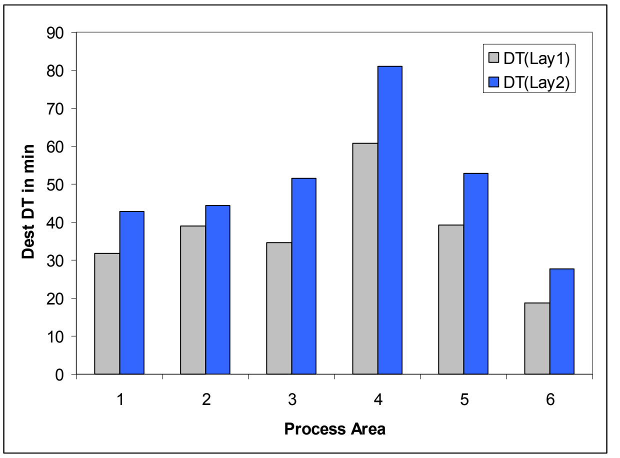

Figure 7: Destination Delivery Time DT $_{dest}$ for each Process Area - Layout 1 vs. Layout 2

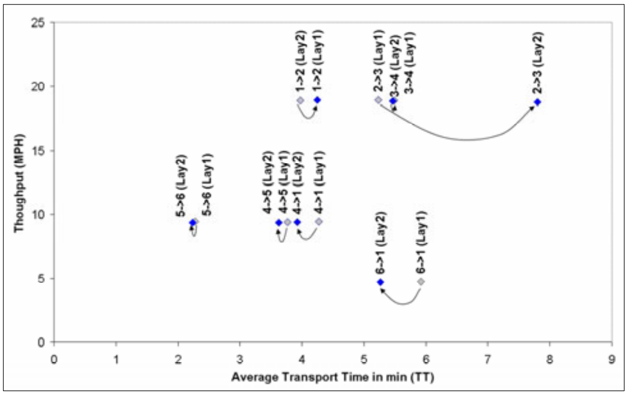

Figure 8: Process Area to Process Area Transportation Time TT(src,dest) - Layout 1 vs. Layout 2

Figure 9 shows the grouping of the transportation times by process areas. $TT_{dest}(3)$ is much higher for Layout 2 compared to Layout 1 whereas a slight decrease in transportation times is observed for all other destination process

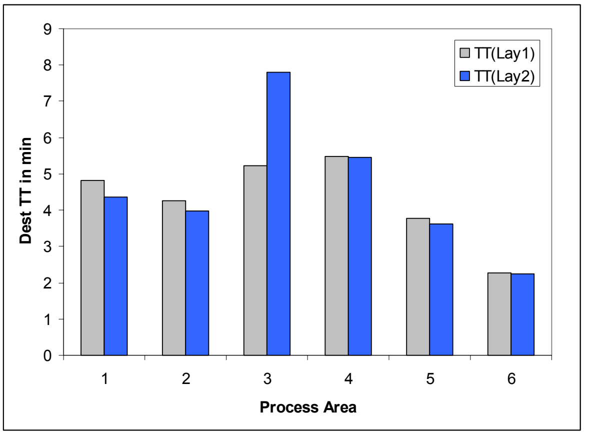

Figure 9: Destination Transportation Time TT_{dest} for each Process Area (Layout 1 vs. Layout 2)

areas. Though, this decrease cannot compensate the worse performance of deliveries to process area 3.

Finally, we perform the congestion analysis in order to investigate whether lower traffic is observed for the changed Layout 2, especially queuing at PPort2. Figure 10 depicts the congestion index for all active node-to-node connections. $C_{TTindex}$ for the connections N2→PPort3, N4→N8, N4→PPort1, N4→PPort4 show high values for Layout 1. N3→N6 is the path with the parking lot which generates a high flow factor due to queuing at the parking lot. As sensitive connections with high TT/Moves gradient, N2→N1 and N9→N8 is observed. In these connections a slight increase of the move rate results most probably in a very strong increase of the congestion index and consequently in TT and DT.

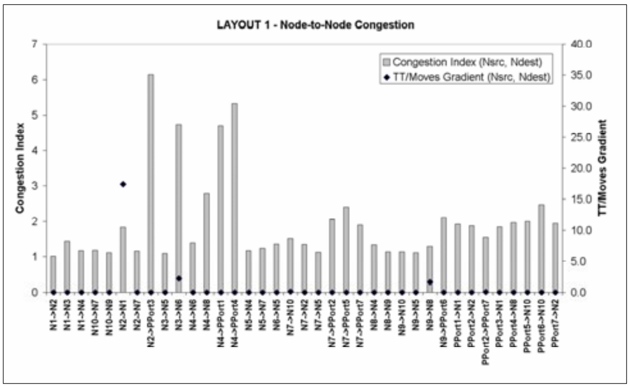

Figure 10: $C_{TTindex}$ and TT/Moves Gradient for Layout 1

The node-to-node congestion index and the TT/Moves gradient for Layout 2 is shown in Figure 11. The CTTindex for N2->PPort3 is unimproved compared to Layout 1. N4->N8 and N4->PPort1 shows a lower congestion index for Layout 2. However, the N2->N1 connection increased the TT/Moves gradient which means that it is more sensitive for congestion and high traffic with increasing move rates in the connection.

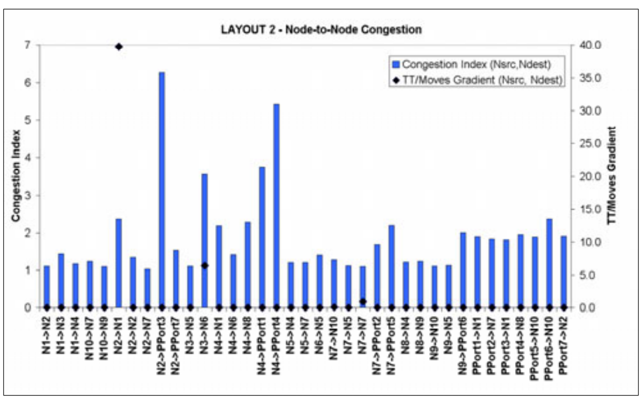

Figure 11: $C_{TTindex}$ and TT/Moves Gradient for Layout 2

In summary, the separation of the PPorts of S2 and S7 in segregated monorail loops shows not a significant improvement of the node-to-node congestion index. The delivery times in all process area to process area connections are increased with the layout change. Especially, process area 3 has a higher delivery time. Consequently, the relative small layout change resulted in a large effect to delivery time performance.

6 CONCLUSION AND OUTLOOK

The efficient comparison of design variants of AMHS can be studied by simulation efficiently when simulation results are calculated based on raw log file data. This approach allows the described three level approach for calculation of throughput and delivery time metrics. Global average metrics over all active move connections are difficult to interpret. More transparency and effects of layout changes with regard to local AMHS performance is obtained from the stocker-to-stocker or node-to-node analysis.

The layout study presented here was on the simulation model derived AMHS installations and installations used for wafer fabs with 200mm wafer size. We currently work on a layout study for a 300mm wafer fab with a unified AMHS with tool-to-tool transportation capability. This adds an additional level in the metrics calculation, which is the process equipment to process equipment view.

REFERENCES

Beamon B.. 1999. System Reliability and Congestion in a Material Handling System. *Computers&Industrial Engineering No.36: 673–684.*

Cardarelli, G. R., Pelagagge, P.M., and Granito, A.. 1996. Performance Analysis of Automated Interbay Material Handling and Storage Systems for Large Wafer Fab. *Robotics&Computer Integrated Manufacturing Vol.12 No.3: 227–234.*

Glüer, D. 2002. Issues for Automation in Semiconductor Manufacturing. *In: at - Automatisierungstechnik, Volume 50, Issue 01, 28.*

Norman, M. 2002. A Recursive Method for Traffic Management through a Complex Path Network. *Proceedings of the 2002 Winter Simulation Conference, 537-541.*

Sturm, R., Seidelmann, J., Reddig, K. 2003. Simulation-based Planning of Automated Material Handling Systems for IC Manufacturing Industries. *wt Werkstattstechnik online Vol. 93 (2003) Issue 3.*

Sturm, R., Reddig, K. 2002. Simulation of AMHS in Transient Manufacturing Scenarios. *Proceedings of Brooks Automation Symposium 2002.*

AUTHOR BIOGRAPHIES

ROLAND STURM is a project manager and scientist at the Fraunhofer Society, Institute of Manufacturing Engineering and Automation (IPA) in the Department of Cleanroom Manufacturing. He holds an M.S. degree in industrial engineering from the University of Erlangen. He works in industrial and research projects in the field of factory modeling and simulation for semiconductor industry with focus on AMHS, equipment capacity and layout planning as well as scheduling and dispatching. His email address is <sturm@ipa.fhg.de> .

JOACHIM SEIDELMANN is a group manager at the Fraunhofer Society, Institute of Manufacturing Engineering and Automation (IPA) in the Department of Cleanroom Manufacturing. He received an M.S. degree in mechanical engineering at the University of Stuttgart. His email address is <seidelmann@ipa.fhg.de>.

JOHANN DORNER obtained his diploma degree in Chemical Engineering at Stuttgart University in 1987. Since 1987 he has been a scientist at the Fraunhofer Society, Institute of Manufacturing Engineering and Automation (IPA) and since 1990 he is the head of the Department Cleanroom Manufacturing and obtained his Ph.D. in 1998. He is working with 40 scientists/engineers and 50 assistant scientists in the area of system & equipment development and production optimization. System & equipment development contains tasks like development, construction, optimization and qualification of equipment and components for clean and high precision demands. Work in the area of production optimization are characterized by detection of potentials for cost reduction and implementation of manufacturing automation methods and systems. Johann Dorner is also active in national and international committees especially for the semiconductor industry. His email address is <dorner@ipa.fhg.de>.

KEVIN REDDIG is a scientist at the Fraunhofer Society, Institute of Manufacturing Engineering and Automation (IPA) in the Department of Cleanroom Manufacturing. He received an M.S. degree in industrial engineering at the University of Karlsruhe. His email address is <reddig@ipa.fhg.de>.