

AN ANALYTICAL MODEL OF VEHICLE-BASED AUTOMATED MATERIAL HANDLING SYSTEMS IN SEMICONDUCTOR FABRS

Dima Nazzal

Industrial Engineering and Management Systems
University of Central Florida
Orlando, FL, U.S.A.

Leon F. McGinnis

School of Industrial & Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332, U.S.A.

ABSTRACT

This research explores analytical models useful in the design of vehicle-based Automated Material Handling Systems (AMHS) to support semiconductor manufacturing. The objective is to correctly estimate the throughput and move request delay. This analysis proposes a computationally effective analytical approach to multi-vehicle AMHS performance modeling for a simple closed loop. A probabilistic model is developed, based on a detailed description of AMHS operations, and the system is analyzed as an extended Markov chain. The model tracks the operations of one vehicle on the closed-loop considering the possibility of vehicle-blocking. This analysis provides the essential parameters such as the blocking probabilities in order to estimate the performance measures. A numerical example is analyzed and simulated using Automod to demonstrate and validate the queuing model.

1 INTRODUCTION

Within the past decade, the semiconductor industry has transitioned from 200mm to 300mm wafer fabrication. The shift to 300mm wafers is a challenging and expensive transition. Full factory automation, required to maximize the productivity of the capital, to satisfy ergonomic restrictions caused by the weight and volume of 300mm wafer lot carriers, and to reduce the particle contamination and vibrational shocks on the wafers (Nadoli and Pillai 1994), depends fundamentally on automated material handling systems (AMHS) as the medium through which factory operations control is implemented. AMHS performance is critical to achieving the planned return on investment (ROI) for the total capital investment. In other words, it is not AMHS cost, it is the AMHS impact on performance that is crucial. AMHS may introduce lot delays or cause tool idle time by failing to move lots in the planned and allotted time. In fact, according to the International Technology Roadmap for Semiconductors (ITRS Report 2005), the key focus areas

and issues for 2005 and beyond is to increase throughput for AMHS, reduce average delivery times, and improve its reliability.

Contemporary simulation technology can produce accurate assessments of fab production performance, including the contribution by the AMHS. However, the corresponding simulation models are both expensive and time-consuming to construct, and require long execution times to produce statistically valid estimates. These attributes render simulation ineffective as a decision support tool in the early phase of system design, where system configurations are likely to change often.

A typical 300mm AMHS has a spine layout, illustrated in Figure 1, with a central material handling spine and loops branching on both sides to serve production equipment. There are two distinct operating scenarios: (1) the spine and the loops are decoupled, and vehicles are dedicated to the spine or to one of the loops; or (2) the spine and the loops are integrated, and vehicles may move freely between them (Pillai et al. 1999). In this research, we consider only the first case, which means a wafer lot moving from a tool on one loop to a tool on a different loop must travel through the main spine, and will use three different vehicles, one in each loop and one in the spine. Automated storage units, referred to as stockers, are used to provide both temporary buffering for work-in-process and transfer between the bay and spine transport systems.

Because of the space restrictions in the 300mm wafer fab bays, vehicle travel is on a unidirectional closed loop without the ability for vehicles to pass each other, even when a vehicle stops to drop-off/pick-up a lot at the input/output port of a process tool or a stocker. Thus, failure to carefully synchronize vehicle movements on a given loop can lead to significant amounts of vehicle blocking and the possibility of lot delay as well as induced vehicle idle time.

Overhead hoist transport (OHT) system serves the move requests originating from the stocker(s) and the processor tools in the bay. Each machine—either stocker

or tool— has two load ports: an input port where loads are dropped off by the vehicle and an output port where loads are picked up by the vehicle to be delivered to their next destinations. Each port can accommodate one vehicle at a time. We use the term station to refer to the input and output ports of the machines. Thus, a loop serving M machines, denoted by m_i , consist of $s = 2M$ stations. Without loss of generality, we assume that the loops starts at the stocker (m_1), and the vehicles' route is assumed to be $m_1, m_{2,\dots}, m_i, m_{i+1}, \dots, m_M, m_1, m_{2,\dots}$

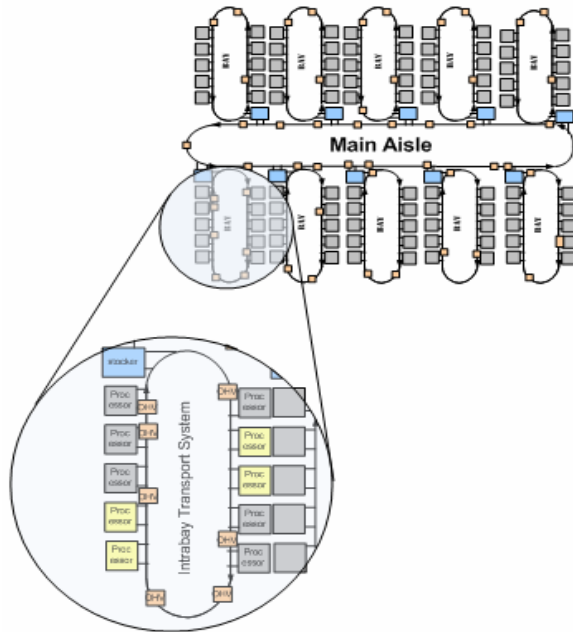


Figure 1: Closed-Loop Unidirectional Interbay and Intra-bay AMHSs

Machine m_i has two stations: the drop-off station s_i^d , and the pick-up station s_i^p . We model an OHT system configured as a simple loop, in which vehicles continuously travel the loop, when a vehicle approaches an m_i , it passes through the drop-off station s_i^d , then travels to the pick-up station s_i^p . Loads are served by vehicles based on the First-Encountered-First-Served (FEFS) rule. FEFS is a decentralized policy, first presented by Bartholdi and Platzman (1989). In FEFS, the vehicles are constantly circulating on the unidirectional loop. When an empty vehicle approaches an m_i , it inspects the output buffer, if there is a load (job) waiting at s_i^p , the vehicle picks it up, which requires time delay l for loading the job and then delivers it to its destination, say machine m_j , visiting machines $m_{i+1}, m_{i+2}, \dots, m_{j-1}$, and finally the load's destination the drop-off station of m_j , denoted by s_j^d . The vehicle does not stop at machines $m_{i+1}, m_{i+2}, \dots, m_{j-1}$ unless it is blocked by other vehicles. If the output port of s_i^p is empty, the vehicle

travels to s_{i+1}^d , then inspects the output port s_{i+1}^p and so forth until it encounters a waiting load.

In order to estimate the throughput accurately, we need to estimate the blocking delays at each machine— either stocker or tool. The main objective of the models to be developed is to quantify the duration of this type of delay as a function of the layout of the transportation system, the demand rates, the speed of the vehicles and the number of vehicles circulating the loop. In this paper, a queuing network type model is developed, based on a detailed description of AMHS operations, and the model is analyzed as an extended Markov chain. With this approach we are able to estimate both AMHS throughput and move request delays.

2 LITERATURE REVIEW

2.1 Models of 300mm Material Handling

Throughout the literature, the importance of AMHS in 300mm wafer fabs has been repeatedly addressed, and research in this area can be broadly categorized into: (1) design optimization, targeting the guide path network layout design and calculating a feasible fleet size, and (2) performance evaluation of various AMHS methods or different AMHS configurations via simulation modeling.

In the area of design optimization, Peters and Yang (1997) propose a network flow formulation to determine the number and location of shortcuts for the interbay transport system in a spine layout fab. The objective function minimizes the tradeoff between the increase in shortcut construction cost and the decrease in material handling costs. Ting and Tanchoco (2000) propose an analytical procedure to construct a unidirectional circular layout for the interbay system in 300mm fabs under the assumption that not all the stockers can be connected by a simple loop. Steele (2002) proposes an algorithm to estimate the performance of an AMHS during the design process. Each design alternative is modeled as a network of nodes. The algorithm estimates the minimum number of vehicles required to deliver the required number of wafer lots and the average delivery times between each pair of source and destination nodes. The algorithm assumes infinite capacity queues and thus does not consider blocking.

To date, discrete event simulation has been the only methodology shown to give reliable estimates of fab-level AMHS performance. At the system design stage, however, large scale, high-fidelity simulation models are not a practical approach, because they take too long to develop, and require multiple lengthy executions to produce statistically valid estimates. Different approaches have been taken to overcome this problem. For example, Mackulak et al. (1998) propose developing a generic model that can be re-configured according to the specific problem at hand, thereby reducing the model building time. Gaxiola and

Mackulak (1999) describe the use of simple deterministic calculations in situations where the process requirements have not yet stabilized.

Pillai et al. (1999) discuss the issue of linking the interbay and intrabay tracks for a 300-mm fab layout. Rust et al. (2002) and Mackulak and Savory (2001) investigate the same problem through simulation by focusing on the impact of this decision on several AMHS performance measures.

Lin et al. (2003) propose using four different vehicle types to carry out the transport tasks from tool to tool. Type A vehicles move in an intrabay system and deliver the lots within the bay. Type B vehicles carry lots between the stockers. Type C vehicles carry lots from a tool in any bay to a stocker in the lot's destination bay. Type D vehicles move lots from a tool in any bay to a tool in any other bay. Three different transport methods using combinations of the four vehicle types were examined.

The conclusions in most of the studies depend on the specifications of the fab being modeled, and thus do not constitute generic design guidelines. At the system design stage, therefore, large scale, high-fidelity simulation models are not a practical approach and system designers are limited in the range of alternatives they can expect to evaluate in detail.

2.2 Analytical Models of Material Handling Systems

Analytical models of AMHS are usually based either on deterministic optimization models or queuing models. The former fails to capture queuing in the system which is essential to accurately estimate the key performance measures. Often, in analytical factory modeling, the material handling system is modeled by defining a "virtual" workstation between the processing tools in a product's routing. The delay associated with material handling is approximated by the processing time on this virtual workstation, which has a capacity equal to the number of vehicles available. This approach is appealing because it exploits well-understood queuing models. However, it has some inherent weaknesses. First, it assumes that the response time of the AMHS to a move request does not depend on the location of the load, nor on the vehicle distribution across the network. Second, it fails to capture the impact of vehicle-to-vehicle blocking, which, by consuming some of the available vehicle time, will degrade the capacity of the AMHS. Vis (2004) provides a survey of work in this area.

Johnson (2001), Johnson and Brandeau (1994, 1995), and Kobza et al. (1998) analyze AMHS using M/G/c queuing models; these models give good approximations provided vehicle assignments are based on a First Come First Served (FCFS) discipline. However, queuing results deviate considerably from simulation results when the vehicle dispatching is system state-dependent, such as Nearest Vehicle Rule (NVR). In Johnson (2001), a queuing model is

used to estimate the performance of a multi-vehicle AMHS with NVR Dispatching. Johnson first develops an approximation for the distribution of the empty vehicles among the stations, then uses an M/G/c model to estimate the waiting time of loads. The latter results tend to be inaccurate because of the assumption of state-independent service time.

Curry et al. (2003) propose a more accurate service-dependent queuing network model that generates approximations that are close to the simulation results but the time to solve the analytic model grows exponentially with the number of vehicles.

Bakkalbasi (1990) develops analytic models to approximate the empty vehicle travel times for the following dispatching rules: FCFS, Closest Load First, Closest Load with Time Priority, and Furthest Load First. Srinivasan et al. (1994) propose a single-vehicle queuing model to estimate the throughput of the vehicle where the vehicle dispatching to move request is based on a modification of the FCFS rule.

Hodgson et al. (1987) have attempted to model single-vehicle systems using Markov decision processes. Due to the large number of states in even a relatively simple AGVS, several constraints were applied to make the Semi-Markov problem tractable.

In Bozer et al. (1994), the throughput approximation is used to estimate the waiting time of move requests at each station; their estimates are quite close to the simulation results. The authors propose an extension of their model to multi-vehicle systems by adjusting the travel times assuming that an AMHS that has K vehicles can be replaced by a single device that travels K times faster. Results indicate good throughput estimates but significant errors in waiting time estimates because congestion and blocking delays are not modeled.

Bartholdi and Platzman (1989) propose the First Encountered First Served (FEFS) decentralized policy for closed loop vehicle-based system. Bozer and Srinivasan (1991) approximate the throughput capacity of single-vehicle system operating under the FEFS policy.

Roeder et al. (2004) propose a simulation of a simplified closed queuing network to model intrabay AMHS in semiconductor manufacturing. The approach has fewer data requirements than an explicit detailed simulation model of the system. The authors use an information taxonomy to quantify the differences between the explicit AMHS simulation and the queuing network approximation. The approximation captures the movement of the vehicles, interaction of the vehicles with the machine loadports, and processing of lots at machines. Vehicle blocking, however, is not modeled.

In summary, analytical models developed under the assumption that loadports have ample capacity for vehicles ignore vehicle-blocking in multi-vehicle systems, a critical characteristic of the system we study, where there are no

offline-docking locations (spurs) and blocking of vehicles is not only possible but very likely. Modeling the AMHS as a single server oversimplifies the system because it assumes that every move request has the same response and travel time. Estimates for AMHS throughput capacity and/or response times based on such models will generate designs that deviate significantly from the actual system.

In an earlier paper (Nazzal and McGinnis 2005), we develop a probabilistic model based on a detailed description of AMHS operations, and analyze the system as an extended Markov chain. The model tracks the operations of all the vehicles on the closed-loop considering the possibility of vehicle-blocking. Steady-state analysis provides estimates of empty-vehicle. The resulting large-scale model provides reasonably accurate estimates; however, it presents some computational challenges. These computational challenges motivated the development of the model we present here that also analyzes the system as an extended Markov chain but with a much reduced state space. The model tracks the movement of a single vehicle in the system with additional assumptions on vehicle-blocking. This reduced-state model offers computationally fast, fairly accurate estimates of the performance measures.

3 MODELING APPROACH

Our objective is to estimate the expected throughput capacity of the AMHS for a given set of input parameters expressed by the throughput requirements, the travel times, the layout of the stations on the AMHS closed loop track, and the fleet size. The analytical model provides estimates of a specific set of AMHS output variables that are essential to calculate the throughput capacity. These output variables are the proportion of time the vehicles spend traveling (empty and loaded), in service (loading and unloading), and being blocked (empty and loaded).

Rather than tracking the location of every vehicle while keeping a record of all the events that change the location and status of the vehicles, we focus on a subset of vehicle operations; operations that vehicles go through only at the drop-off and pick-up stations, eliminating the travel operations that occur on the track segments. Next, we enumerate the vehicle conditions relevant to our analysis. We choose to include the vehicle conditions that identify whether the vehicle has just arrived at the station (empty or loaded), is in service (loading/unloading), or is blocked (empty or loaded).

In Nazzal and McGinnis (2005), we use a transition matrix to track the changes in the locations and the conditions of all the vehicles. Each possible *location-condition-vehicle* combination is identified as a state. The approach proposed here follows similar logic to the previous work but differs in that the Markov chain tracks one vehicle while assuming that there are n vehicles operating in the

AMHS and the move requests are equally distributed among these vehicles.

The transition matrix tracks the changes in the locations and the conditions of one vehicle. In the matrix, each distinct possible *location-condition* pair is identified as a state (in the previous model, a state is identified by *location-condition-vehicle* combination because every vehicle was tracked). The transition between the states is probabilistic and depends on the move requests rate, the number of vehicles, the sequence of the machines on the AMHS loop, and on the possibility of vehicle-blocking. Assumptions on the arrival process of move requests allow us to analyze the transition between the states as a Markov chain.

Some of the transition probabilities are known because they are easily calculated from the given problem parameters (such as the probability that a loaded vehicle will be dropping off its load at some station), while other transition probabilities are only partially determined by the problem parameters, but also influenced by the output variables that we are trying to calculate, specifically the arrival rate of empty-vehicles to stations. We also have a set of transition probabilities that will determine whether the vehicle will get blocked in the next state or not. We did not need these in the previous model because the state specifies where every vehicle is located.

The Markov chain steady-state analysis is combined with necessary conditions that ensure that the AMHS is able to meet the required throughput imposed by the machines. These conditions provide constraints on some of the unknown variables in the Markov chain.

The blocking probabilities that are introduced in this model are estimated by assuming that the probability that a vehicle gets blocked increases linearly with the number of vehicles.

One advantage of this model is that there is no need to create the virtual stations if the travel times and loading times are not equal in order to synchronize the vehicles' movements. However, this implies that unlike the previous model, the transition time between each pair of states is not equal. Therefore, the next step in the new approach is to approximate the transition time between the states, which will depend on the transition probabilities and the given travel and service times.

The Markov chain model combined with the stability conditions, the assumptions on the blocking probabilities and the expected transition times results in a non-conventional Markov chain that has additional sets of unknown variables that were not present in the previous model. However, the new model has a significantly smaller state-space, which does not grow with the number of vehicles in the system. We call this model the reduced-state Extended Markov Chain model. The reduced-state extended model provides a full rank system of equations that can be solved to give a unique set of estimates for the output parameters of the AMHS.

4 THE REDUCED SEMI MARKOV CHAIN MODEL

4.1 Notation

- M : set of tools and stockers in the system.
- m_i : machine i , which could be either the output buffer of the input buffer of a tool or a stocker $m_i \in M$.
- s_i^p : pick-up station of $m_i, i \in M$.
- s_i^d : drop-off station of $m_i, i \in M$.
- $t_{i,i+1}$: time to travel from s_i^p to $s_{i+1}^d; i \in M$.
- t_i : time to travel from s_i^d to $s_i^p; i \in M$.
- l : loading time at $s_i^p; i \in M$.
- n : number of vehicle in the AMHS.
- λ_i : mean arrival rate of move requests from s_i^p .
- q_i : probability that an empty vehicle arriving at a pick-up station s_i^p will find a load waiting for pick-up, $i \in M$.
- r_i : probability that a loaded vehicle arriving at drop-off station s_i^d will drop off its load at $s_i^d, i \in M$.
- p_i^p : the probability that a vehicle is blocked by a vehicle occupying pick-up station $s_i^p, i \in M$.
- p_i^d : the probability that a vehicle is blocked by a vehicle occupying drop-off station $s_i^d, i \in M$.
- R : Set of Markov chain states.
- v_r : the visit ratio to state $r, r \in R$.
- C_r : the expected time between two consecutive visits to state r .
- T_r : the time from the instant the system enters state r until the instant it enters the next state.

We characterize a state by specifying the condition and location of a vehicle. We propose to consider only those points in time where a vehicle is located at a station. A state is defined by three characters:

$$(m_i, i=1,2,\dots,M), (p,d), (e,f,b,k,s).$$

The first and second characters describe, respectively, the machine occupied by the vehicle and the station type (pick-up or drop-off), and the third character specifies the condition of the vehicle (empty, full, blocked/empty, blocked/loaded, and receiving service).

4.2 Transition Probabilities

Let p_{i+1}^d denotes the probability that the vehicle will be blocked by a vehicle at s_{i+1}^d and cannot move from s_i^p to s_{i+1}^d , and let p_i^p denotes the probability that the vehicle will be blocked by s_i^p and cannot move from s_i^d to s_i^p . Let q_i denote the probability that an empty vehicle arriving at a pick-up station s_i^p will find a load waiting, and r_i the probability that a loaded vehicle arriving at drop-off station s_i^d will drop off its load at s_i^d .

The transition matrix \mathbf{R} specifies the movement of the vehicle between the states. The position and type of the vehicle ($e, f, b, k, \text{ or } s$), the possibility of a vehicle blocking its path, and the presence of a load to be picked-up or dropped-off, determine the system transitions. Let $\bar{x} = 1 - x$. In Figure 2 below, consider the transitions from state (i, d, e) , an empty vehicle arriving to a drop-off station s_i^d will move empty to the next pick-up station s_i^p , entering state (i, p, e) if it was not blocked by another vehicle, which happens with probability \bar{p}_i^p . With probability p_i^p , the vehicle gets blocked and transitions to state (i, d, b) . From state (i, p, e) , the empty vehicle will leave empty if there was no load waiting at s_i^p , which happens with probability \bar{q}_i , and also if there is no vehicle at s_{i+1}^d blocking its way, with probability \bar{p}_{i+1}^d , thus, the vehicle moves from state (i, p, e) to state $(i+1, d, e)$ with probability $\bar{q}_i \bar{p}_{i+1}^d$, and to the blocked state (i, p, b) with probability $\bar{q}_i p_{i+1}^d$. However, if the empty vehicle encounters a load at s_i^p , it moves to state (i, p, s) and starts the pick-up process, with probability q_i . Similarly, a loaded vehicle arriving to a drop-off station s_{i+1}^d , a state denoted by $(i+1, d, f)$, will drop-off its load, thus entering state $(i+1, d, s)$ with probability r_{i+1} , or move to states $(i+1, p, f)$ and $(i+1, d, k)$ with probabilities $\bar{r}_i \bar{p}_{i+1}^p$ and $\bar{r}_i p_{i+1}^p$, respectively.

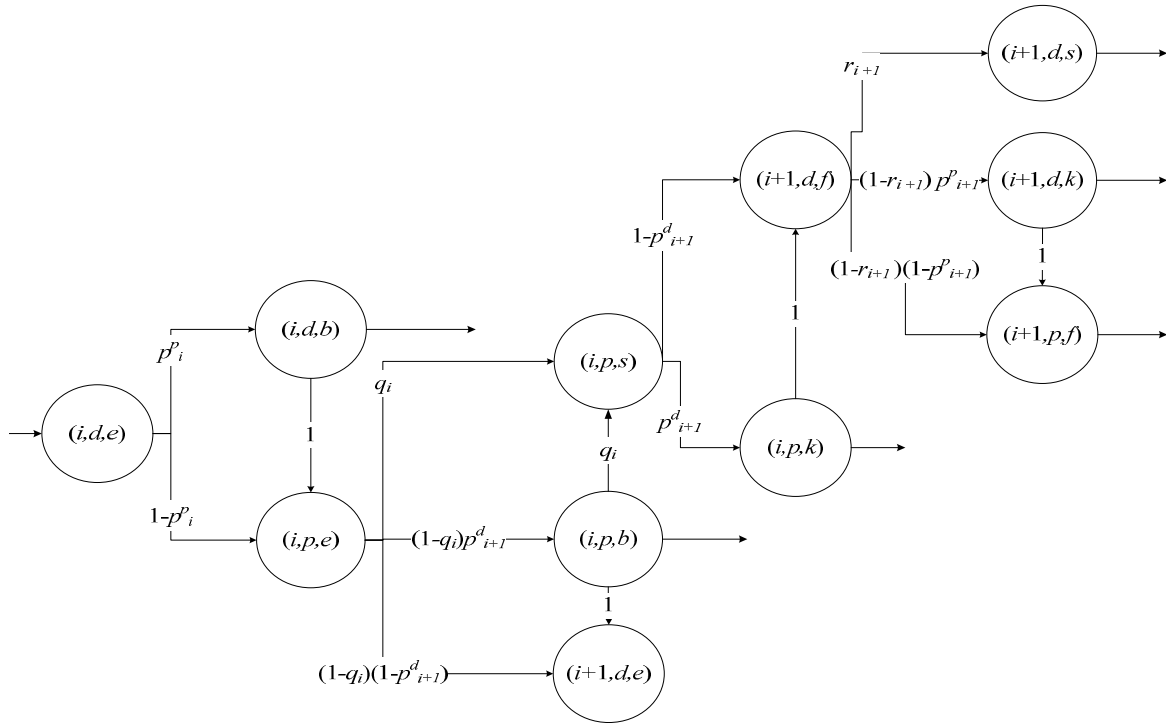


Figure 2: Part of the State Transition Diagram for the Reduced-State Extended Markov Chain

An empty vehicle arriving to a drop-off station will move to the next pick-up station provided that there was no other vehicle blocking its way, and so is the case for a loaded vehicle arriving to a pick-up station. We assume that after a vehicle is blocked, it gets unblocked and moves to the downstream station with probability 1. The justification for this assumption is that a vehicle gets blocked because somewhere downstream another vehicle is in service, this vehicle can be either the one directly downstream from the blocked vehicle or several stations downstream given that there are vehicles occupying all the station in between. The vehicle does not get blocked twice at the same station because the vehicle that was in service has finished its job and will move, unblocking all the vehicles that were waiting behind it.

4.3 The Reduced Markov Chain Steady-State Analysis

Let C_r be the expected time between two consecutive visits to state r , $r=1, \dots, |R|$. Without loss of generality, assume that C is the expected time between two visits to some reference state, say state $(1, p, e)$. Let $v = \{v_r\}$ $r=1, \dots, |R|$, where v_r denotes the visit ratio to state r , which is the number of times the system visits state r between two successive visits to the reference state $(1, p, e)$, by this definition, $v_{(1,p,e)}=1$. For a finite state, positive recurrent Markov Chain, the visit ratios can be uniquely obtained by solving the square system of equations (Ross 2000):

$$\mathbf{R}\mathbf{v} = \mathbf{v} \tag{1}$$

$$v_{(1,p,e)} = 1 \tag{2}$$

The elements in the transition matrix are the transition probabilities between states. Some of these probabilities are unknown, specifically, the load-encountering probabilities $\mathbf{q} = \{q_i\}$, $i=1, \dots, M$, and the blocking probabilities $\mathbf{p}^d = \{p_i^d\}$, $\mathbf{p}^p = \{p_i^p\}$, $i=1, \dots, M$.

The load-drop off probabilities at drop-off stations, $\mathbf{r} = \{r_i\}$, $i=1, \dots, M$ are known. In Nazzal and McGinnis (2005), we demonstrate that they can be easily calculated from the given problem instance parameter using the expression:

$$r_i = \frac{\lambda_i}{\alpha^d}, \forall i \in M \tag{3}$$

Where α^d denotes the rate of loaded vehicles arrivals to drop-off stations. Notice that this parameter is not dependent on the drop-off station since we assume a simple closed loop, and that every load dropped off at s_i^d will be picked up from s_i^p , and therefore the flow of loaded vehicles is equal for every station.

4.4 Stability Conditions

Bozer and Srinivasan (1991) presented stability conditions for a single-vehicle material handling system, we now ex-

tend those conditions for a multi-vehicle system. For a stable system, the expected number of loads delivered by a single vehicle to drop-off station s_i^d in a cycle of length C must equal the number of times a single vehicle enters state (i, d, s) (unloads at s_i^d) in the same period. Let n denote the number of vehicles on the loop, then:

$$v_{(i,d,s)} = \frac{\lambda_i C}{n}, \forall i \in M \quad (4)$$

During a cycle of length C , each vehicle makes $v_{(i,p,e)}$ empty trips to s_i^p . The vehicle picks up a load with probability q_i , and moves to s_{i+1}^d if it does not get blocked. Therefore, during a cycle, the expected number of loads picked up from s_i^p by each vehicle is $q_i v_{(i,p,e)}$. For a stable system, this should equal $\lambda_i C / n$; the number of move requests per vehicle in a time period of length C . Equating these terms we get the following necessary conditions for AMHS stability:

$$q_i = \frac{\lambda_i C}{v_{(i,p,e)} n}, \forall i \in M \quad (5)$$

4.5 Vehicle Blocking Probabilities

A vehicle is blocked when it attempts to move to the downstream station but finds that station occupied by another vehicle. The downstream vehicle could be receiving service, traveling towards that station or also being blocked.

Assuming that blocking occurs when any of the other $n-1$ vehicles are occupying the downstream station and that the blocking probabilities increase linearly with the fleet size, we propose to estimate p_i^d and p_i^p from

$$p_i^d = (n-1) \frac{v_{(i,d,f)} + v_{(i,d,k)} + v_{(i,d,s)} + v_{(i,d,e)} + v_{(i,d,b)}}{\sum_{r \in R} v_r} \quad (6a)$$

$$p_i^p = (n-1) \frac{v_{(i,p,s)} + v_{(i,p,k)} + v_{(i,p,b)}}{\sum_{r \in R} v_r} \quad (6b)$$

4.6 Transition Times

The model tracks the movement of a single vehicle and so it has an advantage over the earlier model developed in Nazzal and McGinnis (2005) from a computational perspective. Here, there is no need for virtual stations in order

to synchronize the vehicles' movement, when the travel times and loading times are not equal. However, this implies that unlike the previous model, the transition times between pairs of states are not equal. We, therefore, need to estimate the expected state transition time $E(T_r)$, defined as the expected time from the instant the vehicle enters state r until the instant it enters the next state.

We do not attempt to show the detailed derivations of the expressions for the expected transition times. But the terms $E(T_r)$, $\forall r \in R$ can be determined based on the transition probabilities. For instance, consider state (i, p, e) , which is an empty vehicle arriving at some pickup station s_i^p , the time that the vehicle spends in this state depends on the probability of encountering a load at s_i^p and on the probability of being blocked at s_i^p by s_{i+1}^d . If the vehicle finds a load at s_i^p with probability q_i , the next state is (i, p, s) ; the transition time in this case is the loading time l , we can follow the same logic to derive the expected transition time from state (i, p, e) if the vehicle does not find a load or gets blocked. Deriving the expected transition times for all the Markov chain states help in developing an expression for the cycle length C by considering the transition time from one state to the next. C was defined as the expected time between two successive visits to the reference state $(1, p, e)$, and can be obtained from:

$$C = \sum_{r=1}^R E(T_r) v_r \quad (7)$$

Combining equation sets (1) through (7), we have a full rank system of equations and we can find the unique solution to the system of equations and calculate the visit ratio to every state, and the blocking probabilities.

5 NUMERICAL EXAMPLE

To evaluate the accuracy of the analytical model, we developed two models: a simulation model that captures the details of an intrabay system of 14 processor tools and one stocker, and an approximate model based on the extended Markov chain model.

The bay serves five products, each has a different route starting from and ending at the stocker. The total arrival rate to the stocker is λ jobs per minute.

We compare the interarrival times of empty vehicles at each pick-up station estimated from the analytic model and a simulation model, at multiple fleet sizes starting with the minimum fleet size that can handle the expected move requests (3 vehicles) up to the maximum fleet size that does not cause AMHS deadlock, 15 vehicles. We used Auto-Mod simulation software to obtain simulated values for this performance measure based on 10 replications of 10

days each; we also made sure that the system reached steady state before we started collecting statistics. In the simulation model, move requests arrive at the stocker according to a Poisson process. However, move request arrivals at processor stations are the result of lot arrival and operation times at the processor station, i.e. we did not force them to follow a Poisson process.

For comparison purposes, we measure the average interarrival time of empty vehicles at all the pick-up stations. This metric is related to the throughput capacity of the AMHS; as the time between two consecutive empty vehicle arrivals to each pick-up station decreases, the AMHS can handle more move requests. The analytical and simulated expected interarrival time of empty vehicles are shown in Table 1. The relative error represents the difference between the analytical and the average obtained from the simulation.

Table 1: Analytical and Simulated Average Expected Time Between Empty Vehicle Interarrival Times (min.)

Fleet size	Tanalytic	Tsimulation	Rel. error
3	6.57	7.01	-0.06
4	5.15	5.74	-0.10
5	4.63	5.24	-0.12
6	4.40	4.86	-0.10
7	4.29	4.59	-0.06
8	4.25	4.45	-0.05
9	4.25	4.35	-0.02
10	4.27	4.33	-0.01
11	4.32	4.40	-0.02
12	4.38	4.55	-0.04
13	4.45	4.83	-0.08
14	4.52	5.19	-0.13
15	4.61	5.70	-0.19

The analytical model performs reasonably well with acceptable error percentages. The accuracy of the model deteriorates at low and high fleet sizes. This is due to the inability of the model to handle the complexity of estimating vehicle-blocking caused by chain-blocking (many vehicles blocking each other). It is expected that chain-blocking occurs when the number of vehicles is large since even when a single vehicle stops to perform service, many vehicles behind it will be blocked. When the number of vehicles is small, chain-blocking is likely to occur because the amount of loading and unloading per vehicle is high and vehicles make frequent stops.

6 CONCLUSIONS AND FUTURE WORK

We presented a reduced-state Discrete Time Markov Chain model that can be used in assessing closed-loop AMHS performance. The model deals with the computational challenges that were observed in the more detailed model

discussed in an earlier publication, Nazzal and McGinnis (2005). The growth of the state space dimensionality is polynomial as opposed to the exponential growth of the previous model.

Experimental comparisons of the model generated results with detailed simulation for small and medium example problems produced acceptable error margins and the results are obtained very quickly. In fact, in simulation modeling, the execution time increases exponentially with the number of vehicles because of the drastic increase in the number of events that the model has to track. The analytical model execution time is a function of the number of stations only.

We are currently working on further validation of the model for more realistic systems, in order to investigate the value of the model in practice to AMHS designers and analysts. There are a number of issues to be explored in further research. First, we will develop expressions for the expected waiting time of loads for a vehicle.

Another issue is the extension of the model to cover more general network configurations. The simple loop structure of the presented model is an issue as recent enhancements to AMHS systems in practice violate this assumption. The extension will require the empty vehicles to be routed probabilistically.

Dispatching policies other than the FEFS policy assumed in this research may be explored. While FEFS is quite simple, it might not be as efficient as other rules that give priority to older or more urgent loads.

We will also explore the possibility of using the AMHS analytic model as a tool used in conjunction with simulation to provide quick performance evaluation of large-scale simulation models such as those used to simulate 300mm wafer fabrication facilities.

ACKNOWLEDGMENTS

We gratefully recognize the generous support by the W. M. Keck Foundation of the Keck Virtual Factory Lab.

REFERENCES

- Bartholdi, J. J. III, and L. K. Platzman. 1989. Decentralized control of automated guided vehicles on a simple loop. *IIE Transactions*, 21: 76-81.
- Bakkalbasi, O. 1990. Flow path network design and layout configuration. Ph.D. Thesis, School of Industrial and Systems Engineering, Georgia Institute of Technology.
- Bozer, Y. A., C. Myeonsig, and M. M. Srinivasan. 1994. Expected waiting times in single-device trip-based material handling systems. *European Journal of Operational Research*, 75: 200-16.
- Bozer Y. A. and M. M. Srinivasan, and C. Myeonsi. 1991. Tandem configurations for automated guided

- vehicle systems and the analysis of single vehicle loops. *IIE Transactions*, 23: 72-82.
- Curry, G. L., B. A. Peters, and M. Lee. 2003. Queueing network model for a class of material-handling systems. *International Journal of Production Research*, 41: 3901-20.
- Gaxiola, G., and G. Mackulak. 1999. Simulation analysis of a semiconductor handling and processing system: process instability can lead to wasted modeling efforts. In *Proceedings of the 31st Annual Summer Computer Simulation Conference*, 137-142.
- Hodgson, T. J., R. E. King, S. K. Monteith, and S. R. Schultz. 1987. Developing control rules for an AGVS using Markov decision processes. *Material Flow*, 4: 85-96.
- ITRS report. 2005. The International Technology Roadmap for Semiconductors: Factory integration., available online via public.itrs.net.
- Johnson, M. E. 2001. Modeling empty vehicle traffic in AGVS design. *International Journal of Production Research*, 39: 2615-33.
- Johnson, M. E., and M. L. Brandeau. 1994. An analytic model for design and analysis of single-vehicle asynchronous material handling systems. *Transportation Science*, 28: 337-53.
- Johnson M. E., and M. L. Brandeau. 1995. Designing multiple-load automated guided vehicle systems for delivering material from a central depot. *Transactions of the ASME. Journal of Engineering for Industry*, 117: 33-41.
- Kobza, J. E., S. Yu-Cheng, and R. J. Reasor. 1998. A stochastic model of empty-vehicle travel time and load request service time in light-traffic material handling systems. *IIE Transactions*, 30: 133-42.
- Lin, J.T., F.K. Wang, and C.K. Wu. 2003. Connecting transport AMHS in a wafer fab. *International Journal of Production Research*, 41: 529-544.
- Mackulak, G., and P. Savory. 2001. A simulation based experiment for comparing AMHS performance in a semiconductor fabrication facility. *IEEE Transactions on Semiconductor Manufacturing*, 14: 273-280.
- Mackulak, G., F. Lawrence, and T. Colvin, T. 1998. Effective simulation model reuse: a case study for AMHS modeling. In *Proceedings of the 1998 Winter Simulation Conference*, 979-984.
- Nadoli, G., and D. Pillai. Simulation in automated material handling systems design for semiconductor manufacturing. In *Proceedings of the 1994 Winter Simulation Conference*, 892-899.
- Nazzal, D., and L. F. McGinnis. 2005. Queueing models of vehicle-based Automated Material Handling Systems in semiconductor fabs. In *Proceedings of the 2005 Winter Simulation Conference*, 2464-2471.
- Peters, B. A., and Yang, T. 1997. Integrated facility layout and material handling system design in semiconductor fabrication facilities. *IEEE Transactions on Semiconductor Manufacturing*, 10: 360-369.
- Pillai, D., T. Quinn, K. Kryder, and D. Charlson. 1999. Integration of 300 mm fab layouts and material handling automation. In *Proceedings of the IEEE/CHMT Ninth International Electronic Manufacturing Technology Symposium*, 23-26.
- Roeder, T., N. Govind, and L. A. Schruben. 2004. A Queueing network approximation of semiconductor automated material handling systems: how much information do we really need? In *Proceedings of the 2004 Winter Simulation Conference*, 1956-1961.
- Ross, S. 2000. *Introduction to Probability Models*, 7th edition, Academic Press.
- Rust, K., R. Wright, and M. Shopbell. 2002. Comparative analysis of 300 mm automated material handling systems. In *Proceeding of the Conference on Modeling and Analysis of Semiconductor Manufacturing (MASM)*, 240-245.
- Srinivasan, M. M., Y. A. Bozer, and C. Myeonsig. 1994. Trip-based material handling systems: throughput capacity analysis. *IIE Transactions*, 26: 70-89.
- Steele, J. 2002. An algorithm for estimating the performance of an automated material handling system for the semiconductor industry. In *Proceeding of the Conference on Modeling and Analysis of Semiconductor Manufacturing (MASM)*, 229-234.
- Ting, J. H., and J. M. Tanchoco. 2000. Unidirectional circular layout for overhead material handling systems. *International Journal of Production Research*, 38: 3913-3936.
- Vis, I. F. A. 2004. Survey of research in the design and control of automated guided vehicle systems. *European Journal of Operational Research*, 170: 677-709.

AUTHOR BIOGRAPHIES

DIMA NAZZAL is an Assistant Professor at the University of Central Florida. She earned her Ph.D. from the School of Industrial and Systems Engineering at the Georgia Institute of Technology. Her research interests are in modeling and design of manufacturing and logistics systems. She can be contacted by email at dnazzal@isye.gatech.edu and her Web address is <http://www.isye.gatech.edu/~dnazzal>.

LEON F. MCGINNIS is Gwaltney Professor of Manufacturing Systems at the Georgia Institute of Technology, and Director of the Product Lifecycle Management Center of Excellence. His primary research interests are in computational models to support design of large-scale complex discrete event logistics systems. His e-mail address is leon.mcginnis@isye.gatech.edu and his Web address is <http://www.isye.gatech.edu/~lfm>.