Neural-Network-Based Delivery Time Estimates for Prioritized 300-mm Automatic Material Handling Operations

Da-Yin Liao, *Member, IEEE,* and Chia-Nan Wang

***Abstract*—This paper deals with lot delivery estimates in a 300-mm automatic material handling system (AMHS), which is composed of several intrabay loops. We adopt a neural network approach to estimate the delivery times for both priority and regular lots. A network model is developed for each intrabay loop. Inputs to the proposed neural network model are the combination of transport requirements, automatic material handling resources, and ratios of priority lots against regular ones. A discrete-event simulation model based on the AMHS in a local 300-mm fab is built. Its outputs are adopted for training the neural network model with the back propagation method. The outputs of the neural network model are the expected delivery times of priority and regular lots in the loop, respectively. For a lot to be transported, its expected delivery time along a potential delivery path is estimated by the summation of all the loop delivery times along the path. A shortest path algorithm is used to find the path with the shortest delivery time among all the possible delivery paths. Numerical experiments based on realistic data from a 300-mm fab indicate that this neural network approach is sound and effective for the prediction of average delivery times. Both the delivery times for priority and regular lots get improved. Specially, for the cases of regular lots, our approach dynamically routes the lots according to the traffic conditions so that the potential blockings in busy loops can be avoided. This neural network approach is applicable to implementing a transport time estimator in dynamic lot dispatching and fab scheduling functions in realizing fully automated 300-mm manufacturing.**

***Index Terms*—AMHS, neural network, prioritized service, 300-mm semiconductor manufacturing.**

I. Introduction

CHIP makers have been challenged with the potential advantages and uncertainties of migration to 300-mm wafer fabrication. The positive side suggests attractive cost benefits, more reliable product quality, and higher productivity. On the down side, there are thousands of unknowns to be clarified or new paradigms yet to be developed before 300-mm manufacturing gets ready for mass production. Highly automated material handling is one of the biggest concerns to the practitioners. Comparing to the operations in 200-mm semiconductor manufacturing, a cost-effective 300-mm fab demands highly automated operations in both processing and material transfer in order to optimize equipment utilization and product cycle times. Due to the increased number of chips from a 300-mm wafer, the required number of wafers is reduced by a factor of 2.25. Therefore, a high-mix 300-mm fab has to suffer from higher varieties of products than a 200-mm fab does. High product mix leads to more frequent process changes and fine tunes on process and metrology equipment. Also, it results in frequent process experiments and inspections as well as frequent pilot or risk production.

In a wafer fab, a lot will be granted high priority, named *Hot Lot* or *Super Hot Lot*, if either it is going to execute several critical operations for experiments or inspections on process conditions, or it was born as a pilot or risk lot for process characterization or design validation before a new product is released to production. Hot lots are very important to both fab operations and product development of IC (Integrated Circuits) designers. Operations of hot lots can be either preemptive against normal operations, or capacity-reserved for no-wait manufacturing. In contemporary 200-mm semiconductor manufacturing, hot lots are specially handled by human operators in order to reduce the transport delay between distant processing equipment. It becomes very challenging to reduce such delays in a 300-mm automatic transport environment. The dynamics of a 300-mm fab are very complicated when incorporating automatic material handling systems (AMHS) into the shop floor.

Manufacturing of high priority lots has well-known significant impacts on production cycle times as well as throughputs of regular production [8], [11], [15]. Such an effect is usually believed to become worse in 300-mm semiconductor manufacturing due to highly automated material handling operations involved. Ehteshami *et al.* [8] conduct object-oriented simulation experiments of a wafer fabrication model to investigate the impact of hot lots on the cycle time of other lots in the system. Their simulation results show that as the proportion of hot lots in the work-in-process (WIP) increases, both the average cycle time and the corresponding standard deviation for all other lot types increase as well. They conclude that hot lots induce either worse services for regular lots or an increase in inventory costs. Fronckowiak *et al.* [11] use a simulation tool, ManSim/X, to analyze the impact for different hot lot distributions for two different products. Narahari and Khan [15] model semiconductor manufacturing systems as re-entrant lines and study the effect of hot lots through an approximate analysis of the re-entrant line

Manuscript received March 31, 2003; revised March 15, 2004. This work is supported in part by National Science Council, R.O.C., under grants of NSC91-2212-E-260-003 and NSC92-2213-E-260-030.

D.-Y. Liao is with the Department of Information Management, National Chi-Nan University, Puli, Nantou 545, Taiwan, R.O.C. (e-mail: dyliao@ncnu.edu.tw).

C.-N. Wang is with the Institute of Industrial Engineering, National Chiao-Tung University, Hsinchu 300, Taiwan, R.O.C. (e-mail: merlinwang90g@nctu.edu.tw).

Digital Object Identifier 10.1109/TSM.2004.831533

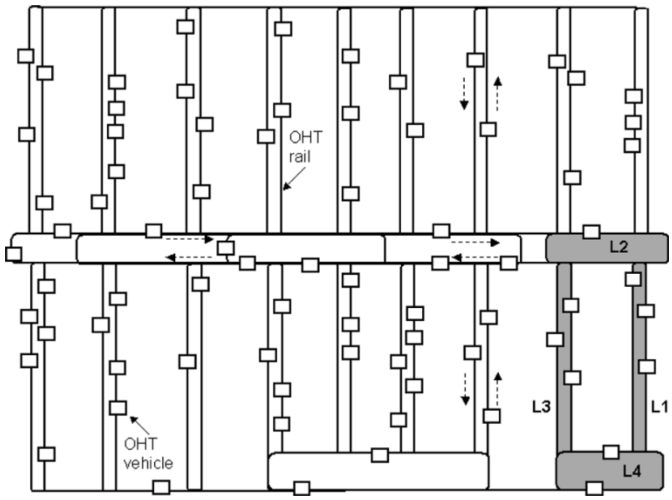

Fig. 1. 300-mm AMHS with OHT loops.

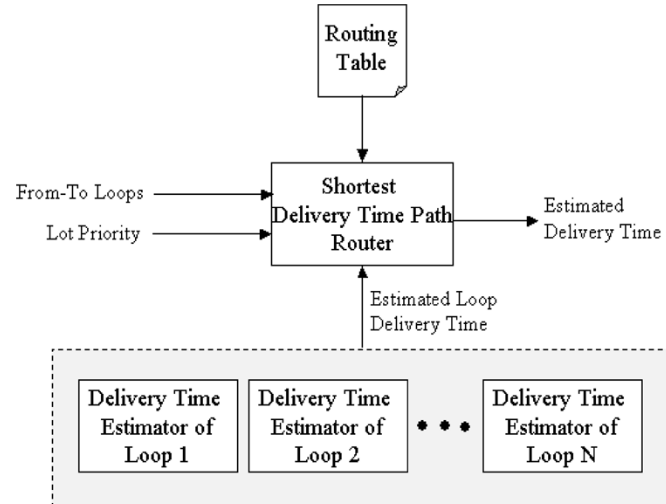

Fig. 2. Loop-to-loop delivery time estimator.

model using mean value analysis (MVA). Their simulation results indicate the significant effect of hot lots on the mean cycle time, variance of cycle time, and throughput rate of regular lots. DeJong and Wu [7] adopt the simulation approach to study the behavior of priority lots in both lot transport and scheduling on machines. Their results reveal that when introducing a relatively small number of priority lots to the AMHS system, significant improvements can be achieved on the delivery performance for the priority lots. For priority lots, the delivery time improvement in an intrabay is better than that in an interbay. Bays with high move volumes and high move request variability will impose big differences between priority and regular lot delivery times.

Among the proposed solutions to 300-mm AMHS implementation, overhead hoist transport (OHT) is one of the most promising technologies [2] in realizing transportation automation in an intrabay, especially good for the operation environment where both automatic and manual carrier transfer operations have to exist simultaneously, like foundry manufacturing [14]. A typical OHT loop is usually designated as a simple directed graph. A 300-mm OHT system can be considered as a combination of dozens of OHT loops, as shown in Fig. 1. Note that tool-to-tool transport is possible in 300-mm AMHS. In this paper, the problems of destination blocking and temporary storage in stockers are handled by fab material control system (MCS) and are out of the scope of this paper. This paper answers the only question of delivery time from one place to another in AMHS.

There have been many research efforts focused on automatic material handling systems in both 300-mm interbay and intrabay [4], [13], [18], [19], [21], [24]. Most of them present the design concept, especially on the effective integration of 300-mm fab layout and AMHS. Cardarelli and Pelagagge [6] develop a simulation tool for design and management optimization of automatic interbay material handling and storage systems for large wafer fabs. Generalized probability density functions fitted on the observation on the monthly input-relative probability in a wafer fab are used as the scenarios to evaluate the dynamics of interbay material handling and storage systems. Liao and Fu [14] discuss the dynamic allocation and dispatching problems in 300-mm AMHS. They suggest an effective OHT dispatch policy, Modified Nearest Job First (MNJF), to achieve high throughputs while reducing the carrier delivery times in a single OHT loop. Bahri *et al.* [5] compare the cost and performance of typical AMHS designs of either partial or fully unified transport with a conventional AMHS that utilizes segregated interbay and intrabay transport systems. Their results indicate that all configurations of direct tool-to-tool transport have shown improvements over their segregated counterpart. Simulations of unified versus segregated AMHS show reduced delivery times of 32% and 66% for normal and hot lots, respectively. In their simulations, lots are not tracked throughout the fab. Instead, lot creation follows a from–to pattern between tool groups and stockers.

This paper deals with the estimation to lot delivery times in 300-mm AMHS. The lot transport time from one location (origin) to another (destination) in a wafer fab is a major input to effective fab dispatching and scheduling functions. Reduction in lot transport times is critical to the production of hot lots whose cycle times are expected to be minimized. However, the determination of lot transport time is usually difficult due to the complicated fab dynamics. Artificial neural networks are one of the most popular techniques in modeling nonlinear and complex dynamic systems [12], [23]. There have been neural network models used for travel time estimation problems in road networks [1], [10]. The computation in a neural network is easy and quickly, as compared to regression methods. This is very helpful to efficient lot delivery time estimation. Efficient and correct estimation of lot delivery time is important to 300-mm shop floor activities like lot dispatching, routing in AMHS, resource reservation for hot lots, and so on.

We adopt a neural network approach for prediction of expected loop-to-loop delivery times of both priority and regular lots, as depicted in Fig. 2. A neural-network model is built for each OHT loop. Inputs for the proposed neural network model are the transport requirements, automatic material handling resources of the loop, and ratio of priority lots in the population. A discrete-event simulation model is built for the automatic material handling functions in a 300-mm fab. Its outputs are used

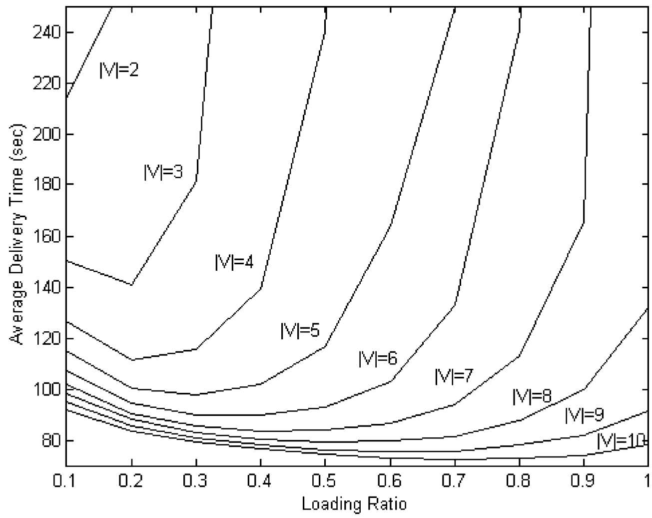

Fig. 3. Average delivery time versus loading ratio.

for training the neural network. We adopt the back propagation method as our training mechanism. The delivery of a lot from its origin to the destination may travel through one or more loops. As lot delivery within a loop is independent of those of the other loops, the average delivery time along a loop-to-loop delivery path can be calculated by adding all the delivery times of each loop along the path. The most efficient delivery of a lot should move along the path with the shortest delivery time. Given the estimates of lot delivery times in each loop, the path with shortest delivery time can be determined by solving an integer programming problem in polynomial computation time. Numerical experiments based on realistic data from a 300-mm fab are conducted to demonstrate the effectiveness of this neural network approach on the prediction of lot delivery times.

The remainder of this paper is organized as follows. Section II presents the neural network model for an OHT loop. Section III describes the discrete-event simulation model of prioritized OHT operations. Numerical experiments of the neural network model are conducted in Section IV. Section V develops the algorithm to generate the shortest delivery time path between the starting and destination loops. Finally, conclusions are made in Section VI with some future research directions.

II. Neural Network Model for an OHT Loop

The dynamics of an OHT intrabay loop are a nonlinear function of transport requirements, number of OHT vehicles, ratio of priority lots against regular ones, number of loadports, and so on. Fig. 3 demonstrates an example of our simulation study on the average delivery time performance against various loading ratios (transport requirements/capacity of design specification) and different OHT numbers, where $|V|$ denotes the number of OHT vehicles.

Computer simulation is one of the techniques that can be successfully applied for analyzing a complex system like AMHS. However, it becomes too time-consuming for practical applications. In this paper, we propose a neural network approach where a delivery time predictor is built based on an artificial neural network model. The neural network model is trained off-line with a discrete-event simulation model where the dynamics of an OHT loop are simulated. The trained neural network model is then adopted as a delivery time estimator which provides an estimate of lot delivery time to fab dispatching and scheduling systems. This neural network delivery time estimator accepts the information of transport loading ratio and priority ratio from fab scheduling/dispatching systems, and the number of OHT vehicles from AMHS. Therefore, instead of being unknown or just with a guess, a better estimation on transport delays in a 300-mm AMHS can be incorporated in generating fab schedules and dispatching decisions. Fig. 4 shows the schematic diagram which demonstrates our neural network approach.

A neural network represents a connection of basic processing units (referred to as neurons) that are capable of processing information in response to external inputs [12], [23]. Neurons in the network are organized in layers. Each neuron of a layer is connected to at least one neuron of another layer in a mesh-like structure. The connection of neurons across layers provides the channel for the transmission of information between neurons. There are three distinct layers characterized in a neural network: the input layer, the hidden layer(s), and the output layer. The development of a neural network model is sometimes with heuristics and the development methods usually depend on the problem itself to be modeled. In this paper, we use a three-layer model for the lot delivery time estimator. The proposed neural network model consists of one input layer, one hidden layer, and one output layer, as shown in Fig. 5. The inputs to the 3-3-2 neural network include loading ratio, priority ratio, and number of vehicles. Its outputs are expected delivery times of priority and regular lots, respectively.

This neural network is first trained off-line to learn the dynamics of an intrabay AMHS which is mimicked by a discrete-event simulation model. During the learning stage, several predefined scenarios are presented to the network as the inputs. For each scenario, the network computes the predicted outputs based on the inputs. The same scenario is also conducted in the AMHS simulation model where the reference outputs are generated. The difference between the computed and the reference outputs is then calculated. The back propagation method is adopted to train the network by updating its weights so that the resulting differences are minimized. To avoid diverging on the network outputs, bounds for the neural network outputs are determined by a threshold limiter. In this paper, we adopt the sigmoid function of the following equation as the threshold limiter of the neural network:

$$f(x) = \frac{1}{1+e^{-x}}. \tag{1}$$

Note that $f(x)$ saturates at 0 and $+1$ when x approaches negative and positive infinity, respectively. However, $f(x)$ holds approximately linear in most of the input space. Such a nonlinearity is commonly adopted for the threshold functions [17].

III. Simulation of Prioritized OHT Operations

A. OHT Intrabay Simulation Model

In order to provide the reference outputs for the neural network model to learn with, a discrete-event simulation model is

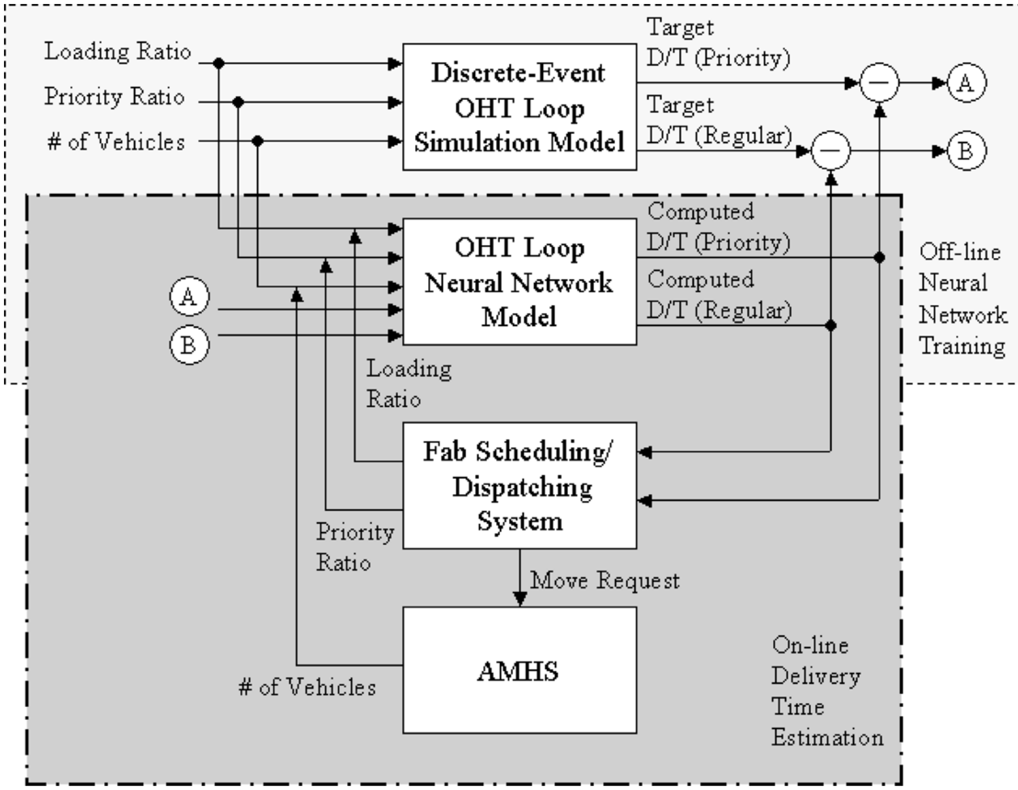

Fig. 4. Neural-network-based delivery time estimates for an OHT loop.

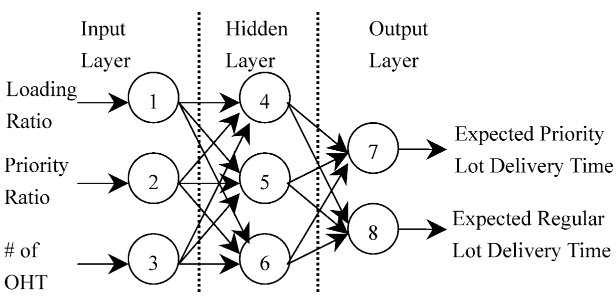

Fig. 5. A three-layer neural network model.

developed, based on realistic data from a local 300-mm wafer fab. This simulation model is implemented with a discrete-event simulation tool—eM-Plant, from Tecnomatix Technologies Ltd. This tool is distinguished by its object-oriented system development capability with characteristics of hierarchy, inheritance and concurrency. Some of the objects are provided in the package, from which users can modify them into user-defined objects for their specific applications. Table I demonstrates some of the objects defined in our 300-mm OHT simulation model.

Fig. 6 depicts an example of the OHT intrabay model in our AMHS simulation system. This OHT intrabay loop is 79.4 meters long and there are two stockers and 22 pieces of equipment. The operating speed of each OHT vehicle is 2 meters per second. It takes 16 s for an OHT vehicle to load or to unload a front-opening unified pod (FOUP) at a loadport. The capacity of design specification in this intrabay loop is 200 moves per hour, where a move is defined as the completion of transport of a lot in the loop. The following assumptions are made in the simulation.

TABLE I
OBJECTS DEFINED IN THE 300-mm OHT SIMULATION MODEL

Item	*Function*	*Defined Object Name*
Event Controller	Control system events	Event Controller
Source	Start of the line	Sources
Special control method	Execute specific actions	IO
Raw spec. data & output	Store system configuration and output results	Performance
Products	Product entity	Regular lot, Hot lot
Loadport	Interface between OHT system and equipment	EQ011,EQ012~EQ223
Stocker Input port	Input gateway of an OHT loop	Stk1In,Stk2In
Stocker output port	Output gateway of an OHT loop	Stk1Out, Stk2Out
Delivery time record	Record deliver time	Deliver trend
OHT	OHT vehicle	OHT
Track	OHT track	Track

A1) As the time of acceleration and deceleration of OHT vehicles are relatively small, they are thus neglected.

A2) There are no failures and maintenance activities on all the vehicles during the simulation horizon. Note that as an OHT loop is a unidirectional circular loop, any failure or maintenance of an OHT vehicle in the loop may cause the entire loop not to function properly. It is meaningless to estimate delivery times in a holdup.

A3) Since we are interested in the effects on the performance of the OHT system, the from-to relationship be-

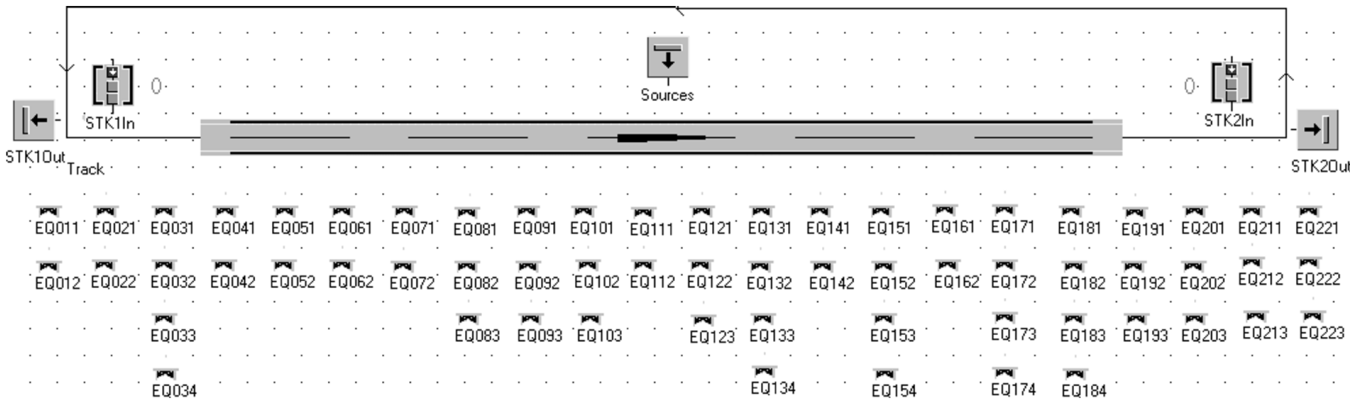

Fig. 6. An OHT intrabay simulation model.

tween two processing units is adopted, instead of considering the whole process flow of a semiconductor product.

A4) The inter-arrival time of transport requests is probabilistic and is assumed to be of exponential distribution.

A5) Unidirectional loadports are adopted. That is, a loadport is either for loading or for unloading. Once a lot is loaded in a loadport, it is transferred immediately. We assume no occupation in a loading loadport. On the other hand, once a lot is transferred to a loadport for unloading, there will be no further transport requests to this loadport before this lot completes its unloading.

B. Prioritized OHT Dispatching

OHT dispatching deals with the assignment of a lot to an empty OHT vehicle. The dispatched OHT is reserved to the lot once it is dispatched, and becomes empty again after completing the transport of this lot. Our objective of OHT dispatching is to minimize lot delivery times in both their mean and variance. A lot with higher priority should dominate those lots with lower priority. Among a given set of transport lots, an empty OHT will be reserved first for the lot with the highest priority. Observing the empirical human operations for carrying high priority products and considering the effect and limitation from OHT transportation, we propose the following OHT dispatching rule to expedite the movement of the highest priority lot in order to avoid any possible blocking due to the transportation of regular lots.

Preemptive Highest Priority Job First (PHP) Policy: Given a set of lots ready for and waiting to be transferred by OHT vehicles in an OHT loop, an empty OHT is dispatched to the lot of the highest priority. The transportation of this OHT vehicle is preemptive. That is, once an empty OHT is dispatched to the highest priority lot, any other ongoing and preemptive transport operations that may block the transportation of this OHT vehicle in the loop will be suspended until this highest priority lot completes its operations in the loop.

In order to study the impact due to this prioritized rule, a rule without differentiating lot priorities, Nearest Job First (NJF) rule, is adopted for comparisons. The NJF rule dispatches a vehicle to its nearest waiting lot. It utilizes the straightforward idea of *first meet*, *first serve* and has been suggested as a good dispatching rule in many automatic guided vehicles (AGV) applications [9], [22].

C. Simulation Model Verification

The developed simulation model is first verified by unit testing. Input and output interfaces, local data structure, and boundary conditions of each defined object in Table I are tested to ensure the information properly flowing into and out of each object under test. Integration testing is then followed to verify the correct movements of OHT vehicles and loading/unloading operations of lots. Logic accuracy of dispatching rules is also tested in the integration testing. We conduct both the unit and integration testing by using very simple examples and tracing the activities of lots and OHT vehicles step by step. Field data with only regular lots from a local 300-mm fab are used for our validation testing to validate the correctness of delivery time performance in the simulation model.

D. Numerical Experiments of the OHT Intrabay Simulation Model

Observing the dynamics of an OHT intrabay system, we found that there are three dominating factors that affect the performance of the OHT system. They are loading ratio, population of priority lots, and number of OHT vehicles in the OHT intrabay loop. These factors are considered in our design of experiments. In order to highlight the effect of the prioritized rules on resource contention, we designate systems with heavy loads. Two loading ratios, 100% and 90% of the design specification, are used in the simulation. As the increasing high-priority population will impose long time delays on the regular jobs drastically, two distributions of high-priority lots, 2% and 8%, are selected for the tests. In addition, if we increase the number of OHT vehicles, the resulting system performance usually improves. In the simulation study, we consider two configurations of the OHT numbers, 4 and 6 OHT vehicles in the loop, respectively. Eight simulation experiments are then conducted based on the scenarios for these three controls. Lot delivery time, the time from the creation of a transport request for a lot to its completion, is considered as the performance measure. For each simulation experiment, we take the statistics of 1000 runs with different initial values of random number

TABLE II
EXPERIMENT RESULTS IN LOT DELIVERY TIME (IN SECONDS)

System Configuration			NJF		PHP			
					Priority		Regular	
Loading Ratio	Priority Ratio	# of OHTs	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
90%	2%	4	97	10	73	27	96	20
90%	8%	4	96	10	74	35	106	21
100%	2%	4	102	10	74	23	105	18
100%	8%	4	102	10	73	33	112	20
90%	2%	6	85	9	71	34	87	17
90%	8%	6	87	9	74	36	89	19
100%	2%	6	86	9	74	37	87	18
100%	8%	6	87	9	74	38	90	21
Average			93	10	73	33	97	19

seeds. The simulation horizon is set to one day long in time unit of seconds. That is, there are 4800 transport requests during the simulation horizon. All the statistics are calculated only after 0.5 simulation hour of each simulation run. Table II demonstrates the simulation results.

In order to clarify the differences between the PHP and NJF rules on the delivery time performance of priority and regular lots, we adopt the analysis of variance procedures to test whether these results are relatively similar or homogeneous. We have the null hypothesis

H_0: No difference in the delivery times for priority lots;

versus the alternative hypothesis

H_a: At least two of the delivery times differ.

The t-value for testing the cases of priority lots is 13.49, which implies that the null hypothesis H_0 is rejected. That is, the performance results from PHP and NJF are statistically different for priority lots. On the other hand, the t-value for testing the cases of regular lots is -1.57, which implies we cannot reject the null hypothesis, H_0: no difference in the delivery times for regular lots. Therefore, there is no significant difference between PHP and NJF for regular lots.

In all of the test scenarios, the PHP rule does expedite the delivery times of priority lots. However, longer delivery times of regular lots are incurred in both mean and standard deviation. In average, the mean delivery time of priority lots can be reduced by 22% by the PHP rule. The PHP rule is thus considered effective in reducing lot delivery times of priority lots with acceptable time delay on regular lots with in average 5% of increase. In these scenarios, the average theoretical lot delivery time ($=$ loading time $+$ travel time $+$ unloading time, the ultimate delivery time without suffering any transport delay) is 51.8 s. That is, the average transport delay is $73 - 51.8 = 21.2$ s for the priority lots applied with the PHP rule. These transport delays are caused by the waiting times before an empty OHT picks up the priority lot.

IV. NUMERICAL EXPERIMENTS OF THE NEURAL NETWORK MODEL

In order to generate reference targets to train the 3-3-2 neural network, more simulation studies are conducted with the number of OHT vehicles ranging from 3 to 6, with loading ratios ranging from 90% to 100%, and with priority ratio ranging from 2% to 10%, respectively. The simulation results of average lot delivery times are listed in Table III. Among these scenarios, results of 100 tested scenarios are then selected to serve as the target outputs for the neural network training process. The back propagation method is conducted with sigmoid activation for 50 000 cycles. Test patterns of 40 scenarios in Table III, different from those selected for training, are used to test the effectiveness of the trained neural network model. We adopt the package of QNET [20] to develop the neural network model. Table IV presents the parameters used for training the neural network. The resultant weights of the trained neurons are listed in Table V. Both the root mean square (RMS) errors of the sets of training and testing patterns converge after 16 000 iterations, as depicted in Fig. 7 and 8. The comparisons between the network outputs and the training targets are shown in Fig. 9. Note that both measures of network outputs and training targets have been normalized by the formula of $(\text{data value} - \text{the smallest})/(\text{the largest} - \text{the smallest})$. All these results indicate that the neural network model provides accurate estimates of the simulation outputs.

TABLE III
AVERAGE LOT DELIVERY TIMES IN THE SIMULATION RESULTS

Configuration		**3 OHTs**		**4 OHTs**		**5 OHTs**		**6 OHTs**	
Loading	**Priority Ratio**	**Regular**	**Priority**	**Regular**	**Priority**	**Regular**	**Priority**	**Regular**	**Priority**
90%	**2%**	168	76	96	73	90	68	87	71
	4%	178	76	102	72	91	71	87	70
	6%	196	77	102	73	92	73	88	71
	8%	220	80	106	74	92	70	89	74
	10%	247	81	109	75	95	73	91	71
92.5%	**2%**	196	78	101	71	89	69	87	68
	4%	236	79	105	73	90	73	88	70
	6%	249	81	106	72	92	69	90	72
	8%	292	79	109	72	93	70	89	70
	10%	360	81	113	75	95	71	90	71
95%	**2%**	245	77	102	69	90	70	86	68
	4%	250	77	105	73	91	71	89	69
	6%	385	80	110	73	93	71	90	69
	8%	X	81	111	73	95	72	91	72
	10%	X	83	112	74	99	73	92	72
97.5%	**2%**	320	75	104	72	91	70	86	71
	4%	402	78	111	73	92	71	89	71
	6%	X	85	112	73	95	72	89	71
	8%	X	85	116	74	96	70	91	71
	10%	X	83	123	76	97	73	92	72
100%	**2%**	X	87	105	74	92	71	87	74
	4%	X	81	111	73	93	72	89	70
	6%	X	83	112	73	94	72	90	70
	8%	X	81	112	73	97	72	90	74
	10%	X	83	124	74	100	72	92	72

Remark: X indicates that the case diverges.

V. NEURAL-NETWORK-BASED DELIVERY TIME ESTIMATOR

As the delivery of lots within an OHT intrabay loop is independent of those of the other loops, the expected delivery time along the delivery path of a lot is the summation of the delivery times of each loop in the path. Assume that there are L OHT intrabay loops in a fab. Let w_i and w_j denote the estimate of lot delivery times in loops i and j, respectively. Both w_i and w_j are

TABLE IV
TRAINING PARAMETERS OF THE NEURAL NETWORK MODEL

Function/Parameter	Value
Training Method	Back Propagation
Number of Layers	3
Input Layer Nodes	3
Input Layer Transfer Function	Linear
Hidden Layer 1 Nodes	3
Hidden Layer 1 Transfer Function	Sigmoid
Output Layer Nodes	2
Output Layer Transfer Function	Sigmoid
Iterations	50000
Training Patterns	100
Testing Patterns	40
Training RMS Error	0.0386100
Training Correlation	0.974929
Testing RMS Error	0.039220
Testing Correlation	0.983577
Learn Rate	0.150000
Momentum Factor	0.850000

TABLE V
RESULTANT WEIGHTS AFTER TRAINING

Connection Node		Weight
From	To	
1	4	1.16976
1	5	4.59572
1	6	-2.78809
2	4	-6.47795
2	5	-0.21034
2	6	18.76053
3	4	-1.74859
3	5	-5.36069
3	6	3.47815
4	7	0.13508
4	8	2.65252
5	7	-3.61648
5	8	1.54068
6	7	1.88680
6	8	-2.57542

generated by the neural networks of loops i and j, respectively. The path of shortest delivery time, starting from loop s to loop f, can be determined by solving the following integer programming problem (P):

$$\text{(P)} \qquad \underset{\{x_{ij}\}}{\text{Minimize}} \sum_{j=1}^{L} \sum_{i=1, i\neq j}^{L} \left(\frac{1}{2}(w_i + w_j)x_{ij} + \frac{1}{2} w_s x_{sj} + \frac{1}{2} w_f x_{if} \right)$$

subject to

$$\sum_{j=1}^{L} x_{sj} = 1, \tag{2}$$

$$\sum_{i=1}^{L} x_{if} = 1, \tag{3}$$

$$\sum_{i=1}^{L} x_{ik} = \sum_{j=1}^{L} x_{kj}, \quad \forall k \neq s \text{ and } f, \tag{4}$$

$$x_{ij} = 1 \text{ or } 0, \quad \forall i \neq j, \tag{5}$$

$$x_{ij} = 0, \quad \text{if } x_{ij} \notin \text{R}, \tag{6}$$

where x_{ij} is the decision variable indicating the selection of the directly-connected path from loop i to loop j, and R is the collection of directly-connected paths which are defined in the routing table of Fig. 2. Problem (P) can be solved by integer programming techniques in polynomial computational time [16]. We adopt Dijkstra's shortest path algorithm [3] to solve Problem (P). We implement the Shortest Delivery Time Path (SDTP) algorithm in C language, with its pseudo code listed in Table VI.

In order to demonstrate the ability of finding the path of the shortest delivery time, consider the following testing case where lots are transported among loops in the shaded area of Fig. 1. We designate 1,000 lots (40 priority and 960 regular lots) to be transported from L1 to L3. There are two possible paths for

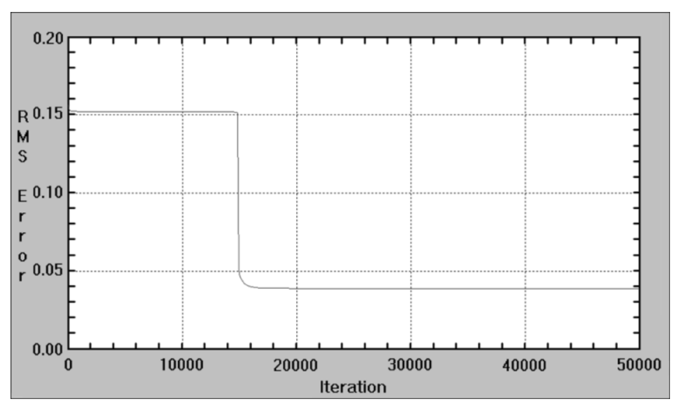

Fig. 7. Training set RMS error.

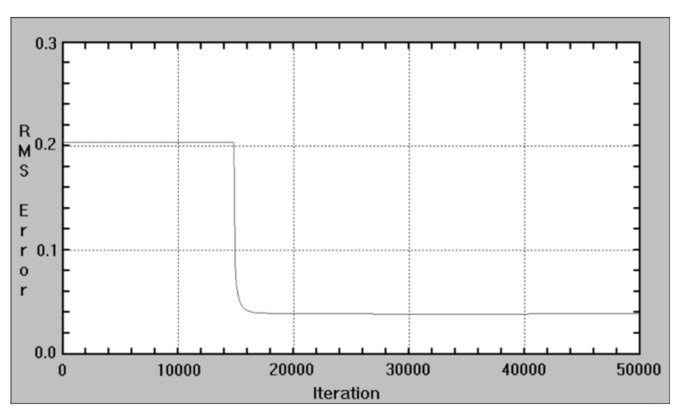

Fig. 8. Testing set RMS error.

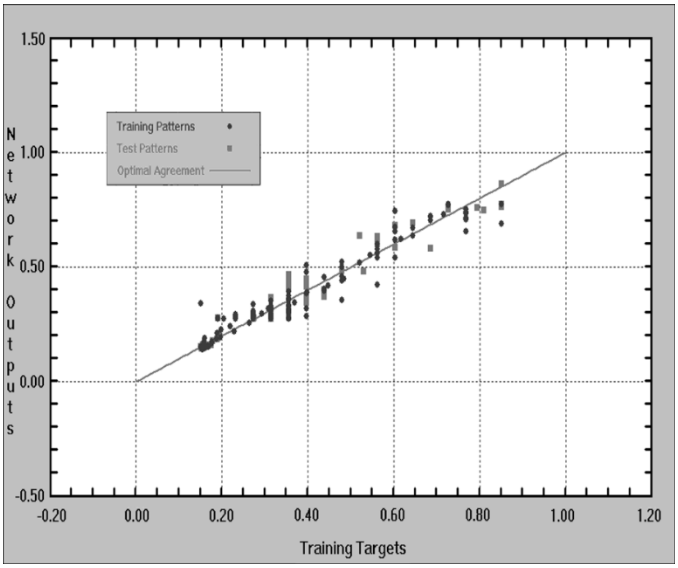

Fig. 9. Comparisons between target and test patterns.

TABLE VI
SHORTEST DELIVERY TIME PATH (SDTP) ALGORITHM

```
SDTP(RoutingTable, LoopDeliveryTime, StartLoop, FinishLoop)
// Assumption: There exists at least one path from StartLoop to FinishLoop
// Initialization
    construct a graph G = (V, E), where v ∈ V[G] and e ∈ E[G]: RoutingTable
    for vertices u, v ∈ V[G]
        do  w(u, v) ← 0.5(LoopDeliveryTime of u) + 0.5(LoopDeliveryTime of v)
    for each vertex v ∈ V[G]
        do  d[v] ← ∞
            p[v] ← NIL
    d[StartLoop] ← 0
    R ← ∅
    Q ← V[G]
// Finding the shortest delivery time path from StartLoop to FinishLoop
    while Q ≠ ∅ and FinishLoop ∉ R
        do  u ← GET-MIN(Q)        // Get the vertex with minimum d[v]
            Q ← Q − {u}
            R ← R ∪ {u}
            for each vertex v ∈ Adj[u]
                if  d[u] > d[v] + w(u, v)
                    then  d[v] ← d[u] + w(u, v)
                          p[v] ← u
// Output the path from FinishLoop backward to StartLoop
    v ← FinishLoop
    output v
    while ( p[v] ≠ NIL )
        do  output p[v]
            v ← p[v]
// End of SDTP
```

lots to travel from L1 and L3, namely Path A and Path B, as described below:

$$\text{Path A}: \quad \text{L1} \Rightarrow \text{L2} \Rightarrow \text{L3} \;\; \text{and} \;\; \text{Path B}: \quad \text{L1} \Rightarrow \text{L4} \Rightarrow \text{L3}.$$

Table VII lists the estimate of delivery time in each loop, which is predicted by the neural network model for each loop. The traffic condition in L2 is designated very busy, while the traffic

TABLE VII
ESTIMATE OF LOOP DELIVERY TIMES FROM THE NEURAL NETWORK ESTIMATOR

		L1	L2	L3	L4
Mean Loop	Priority	71	81	74	70
Delivery Time	Regular	88	275	106	86

TABLE VIII
COMPARISONS OF DELIVERY TIMES BETWEEN THE SDTP AND THE ND RULES

	SDTP		ND (Nearest Depot)	
	mean	variance	mean	variance
Priority	215	130	231	127
Regular	284	68	473	76

in other loops are moderate. Experiments of both the Shortest Delivery Time Path (SDTP) algorithm and the Nearest Depot (ND) rule are conducted. In the Nearest Depot rule, the delivery path is determined by the topologically shortest distance, that is, a lot is routed to the nearest exit to the current loop and to the nearest entry point to the destination loop. In SDTP, the delivery path is selected according to the estimate of the delivery time. Experiment results are shown in Table VIII. It takes less than 0.5 s of the computation time of SDTP in a Pentium-IV 1.8 GHz PC for all the tested cases. The delivery times by SDTP improve for both priority and regular transportation. Specially, for the cases of regular lots, SDTP dynamically routes the lots according to the traffic conditions so that the potential blockings in busy loops can be avoided.

VI. CONCLUSION

A neural-network-based estimator is proposed to predict the average loop-to-loop delivery times of both priority and regular lots. A 3-3-2 neural-network model is built for each AMHS intrabay loop, whose inputs are the transport loading ratio in the design capacity, automated material handling resources, and priority/regular job ratios. A discrete-event simulation model is built for the automatic material handling functions in a 300-mm fab and its outputs are adopted for neural network training. As the delivery of lots within an intrabay loop is independent of those of the other loops, the average delivery time along the loop-to-loop delivery path is the summation of all the delivery times of each loop of the path. The delivery of a lot should move along the path that results in the shortest delivery time. Given the estimate of lot delivery time in each loop, the path with the shortest delivery time can be determined by solving a shortest path algorithm in polynomial computational time. Numerical experiments based on realistic data from a 300-mm fab indicate that this neural network model is sound and effective for the prediction of average delivery times. The developed approach is applicable to the implementation of a transport time estimator for 300-mm fab scheduling and dispatching systems.

Future research directions include the point-to-point delivery time estimate within an OHT loop so that the tool-to-tool delivery time estimate in a fab can be realized. However, due to the complicated dynamics in an automated fab, the estimate of tool-to-tool delivery time may suffer from a large spread of distribution. Estimates from such a wide confidence interval become useless information for fab scheduling and dispatching. As there may be several OHT dispatching rules applied in different intrabay loops in the same time for different transport purposes, the other research direction is to incorporate more dispatching rules, e.g., Longest Waiting Time First (LWT) rule or Equal In-between Distance (EID) rule, into our neural network model so that its estimates will be closer to the performance from the real OHT system.

REFERENCES

[1] J. Anderson and M. Bell, "Travel time estimation in urban road networks," in *Proc. IEEE Conf. Intelligent Transportation Systems*, Nov. 1997, pp. 924–929.

[2] Automated Material Handling System (AMHS) Framework User Requirements Document: Version 1.0, International SEMATECH, 1999.

[3] M. Ataliah, *Algorithms and Theory of Computation Handbook*. Boca Raton, FL: CRC Press LLC, 1999.

[4] U. Bader, J. Dorner, J. Schlieber, T. Kaufmann, and M. Garbers, "Cost efficient future automation and transport concepts," in *Proc. 7th Int. Symp. Semiconductor Manufacturing*, Tokyo, Japan, Oct. 1998, pp. 57–60.

[5] N. Bahri, J. Reiss, and B. Doherty, "Comparison of unified vs. segregated automated material handling," in *Proc. IEEE Int. Symp. Semiconductor Manufacturing*, San Jose, CA, Oct. 2001, pp. 3–6.

[6] G. Cardarelli and P. M. Pelagagge, "Simulation tool for design and management optimization of automated interbay material handling and storage systems for large wafer fab," *IEEE Trans. Semiconduct. Manufact.*, vol. 8, pp. 44–49, Feb. 1995.

[7] C. D. DeJong and S. P. Wu, "Simulating the transport and scheduling of priority lots in semiconductor factories," in *Proc. 2002 Winter Simulation Conf.*, San Diego, CA, Dec. 2002, pp. 1387–1391.

[8] B. Ehteshami, R. B. Petrakian, and P. M. Shabe, "Trade-offs in cycle time management: Hot lots," *IEEE Trans. Semiconduct. Manufact.*, vol. 5, pp. 101–106, May 1992.

[9] P. J. Egbulu and J. M. A. Tanchoco, "Characterization of automatic guided vehicle dispatching rules," *Int. J. Prod. Res.*, vol. 22, no. 3, pp. 359–374, 1984.

[10] L. Fu and L. R. Rilett, "Dynamic O-D travel time estimation using an artificial neural network," in *Proc. Vehicle Information & Navigation Systems, 6th Annu. VINS*, Seattle, WA, July 1995, pp. 236–242.

[11] D. Fronckowiak, A. Peikert, and K. Nishinohara, "Using discrete event simulator to analyze the impact of job priorities on cycle time in semiconductor manufacturing," in *Proc. IEEE/SEMI Advanced Semiconductor Manufacturing Conf. and Workshop*, 1996, pp. 151–155.

[12] F. M. Ham and I. Kostanic, *Principles of Neurocomputing for Science & Engineering*. New York: McGraw Hill, 2001.

[13] R. Kurosaki, T. Shimura, H. Komada, T. Kojima, and Y. Watanabe, "Low cost and short lead time AMHS design using interbay/intrabay diverging and converging method for 300 mm fab," in *Proc. 9th IEEE Int. Symp. Semiconductor Manufacturing*, 2000, pp. 48–51.

[14] D.-Y. Liao and H.-S. Fu, "Dynamic OHT allocation and dispatching in large-scaled 300 mm AMHS management," in *Proc. 2002 IEEE Int. Conf. Robotics and Automation*, Washington, DC, May 2002, pp. 3630–3636.

[15] Y. Narahari and L. M. Khan, "Modeling the effect of hot lots in semiconductor manufacturing systems," *IEEE Trans. Semiconduct. Manufact.*, vol. 10, pp. 185–188, Feb. 1997.

[16] S. G. Nash and A. Sofer, *Linear and Nonlinear Programming*. New York: McGraw Hill, 1996.

[17] J. C. Principe, N. R. Euliano, and W. C. Lefebvre, *Neural and Adaptive Systems*. New York: Wiley, 2000.

[18] D. Pillai, T. Quinn, K. Kryder, and D. Charlson, "Integration of 300 mm fab layouts and material handling automation," in *Proc. 1999 IEEE Int. Symp. Semiconductor Manufacturing*, October 1999, pp. 23–26.

[19] D. Pillai and S. Srinivasan, "Material handling automation—trends, vision, and future plans," in *Proc. IEEE/SEMI Int. Symp. Semiconductor Manufacturing*, 1997, pp. 251–254.

[20] Qnet V2000, Vesta Service, Inc., Winetka, IL*(http://www.qnetv2k.com)*

[21] B. Subramaniam and D. K. Kryder, "Automation challenges in the next generation semiconductor factory," in *Proc. IEEE/SEMI Advanced Semiconductor Manufacturing Conf.*, 1997, pp. 349–355.

[22] J.-H. Ting and J. M. A. Tanchoco, "Unidirectional circular layout for overhead material handling systems," *Int. J. Prod. Res.*, vol. 38, no. 16, pp. 3913–3935, 2000.

[23] L. H. Tsoukalas and R. E. Uhrig, *Fuzzy and Neural Approaches in Engineering*. New York: Wiley, 1997.

[24] J. Weckman, "300 mm fab/AMHS layout challenge: A cookbook approach," in *Proc. 7th IEEE Int. Symp. Semiconductor Manufacturing*, 1998, pp. 61–64.

Da-Yin Liao (S'90–M'91) received the B.S. degree in mechanical engineering and the M.S. and Ph.D. degrees in electrical engineering from National Taiwan University, Taipei, Taiwan, R.O.C., in 1989, 1991, and 1994, respectively.

From 1994 to 1996, he served as a Second Lieutenant in the Chinese Army, Taiwan. He is currently with Department of Information Management, National Chi-Nan University, Taiwan, R.O.C., where he started his professorship from September 2001. Before then, he has worked as a department manager and a senior director in the IT Division of semiconductor foundry and TFT/LCD manufacturing companies for six years. He worked in a 300-mm mass production fab on the design and implementation of computer-integrated manufacturing (CIM) and automatic material handling systems (AMHS) before he joined National Chi-Nan University. He has participated and leaded the development of CIM and AMHS projects in four wafer fabs and one TFT/LCD fab. His current research interests include computer-integrated manufacturing, production scheduling, discrete optimization, and production management of semiconductor manufacturing.

Dr. Liao served as a session chair/co-chair for many international conferences.

Chia-Nan Wang is working toward the Ph.D. degree at National Chiao Tung University, Taiwan, R.O.C.

He works as a director of Newfancy Technology Inc. He has worked as a section manager in manufacturing department and was in charge of the design and implementation of computer-integrated manufacturing (CIM) and automatic material handling systems (AMHS) for Taiwan Semiconductor Manufacturing Company (TSMC) for six years. He also worked as a project manager for evaluation and development of AMHS systems at International SEMATECH (ISMT) for two years. His research interests include production management of semiconductor, technology management, information management and multimedia application.