ABC: A better control for manufacturing

by Jerome M. Kurtzberg
Menachem Levanoni

ABC is a generic methodology to improve the quality of manufacturing. It can optimize operation of a single process or an entire factory to meet or exceed product specifications. ABC is based on three nets which continually interact to model processes and to provide local process control and global product optimization. Significant process variables are identified, evaluated, and ranked according to their contributions to product quality. Process performance, which determines product quality, is characterized by a sensitive parameter, the Q-factor, which is used for local control and for global optimization. Real-time response maps capture process behavior and identify current status, improved operating points, and expected improvement in relation to design targets. ABC continually compensates for off-specification manufacturing steps by feedforward-and-feedback corrective actions which keep the product on target. ABC also evaluates and ranks the effects of non-numeric manufacturing variables, such as specific tools and vendors, on product quality. Total quality control can be achieved by optimizing all variables, both sensor-based and non-numeric, which control the product. Some of ABC's capabilities are demonstrated in a multistep fabrication of a semiconductor capacitor in which the electrical properties of the product are optimized by controlling the individual chemical process steps. ABC's capacity to minimize scrap and rework by compensating for out-of-control conditions is demonstrated in this example. A functional subset of ABC currently exists as a menu-driven tool, implemented in APL2® on VM/CMS for mainframe computers and in the C language for workstation platforms: RS/6000 running under AIX® and PS/2® under OS/2®. ABC is available, in the workstation version, as an IBM Program Offering under the name *QuMAP™—A Better Control*, and is currently used in the semiconductor, pharmaceutical, chemical, and consumer goods industries.

1. Introduction

ABC is a generic methodology that models, monitors, controls, and optimizes product quality in manufacturing from a single process step to an entire factory [1, 2]. The methodology applies to a wide range of manufacturing applications. A functional subset of ABC has been developed for mainframe computers, running on VM/CMS, and is also available on workstations using OS/2® or AIX® as an IBM Program Offering under the name *QuMAP™—A Better Control* [3]. QuMAP is currently in use in the semiconductor, pharmaceutical, chemical, and consumer goods industries.

Previous methods for managing manufacturing processes [4–6] and industrial experiments [7, 8], statistical packages [9, 10], and control charts [11] provided process status information. These approaches, unlike ABC, do not provide comprehensive and systematic process analysis or product quality control, and they do not optimize an entire manufacturing process.

ABC is structured into three interacting nets: the A-net for modeling, the B-net for control, and the C-net for optimization. The A-net models the processes. It tracks process parameters and models the relationships between input and output process variables. The B-net controls processes locally by driving each of them to its target point. The C-net optimizes the entire set of processes. It determines the process target points which provide overall optimization. The C-net compensates for an individual process which does not meet its target by adjusting the remaining process targets for overall improvement.

To manage highly complex manufacturing enterprises, ABC utilizes hierarchical, dynamic modeling with recursive control and optimization. At each hierarchical level, the process is partitioned into logical groups so that treatment is simplified and manageable. Further, ABC minimizes the number of parameters required to represent the manufacturing process, ranks contributions of variables to the process by significance, and automatically detects the existence of unmeasured significant variables. This provides the user with a better understanding of the process and answers the following important questions: *Am I monitoring all relevant parameters? Which parameters are important to my process? Which parameters have little effect on my process?*

ABC employs a unique, highly sensitive measure of product quality for the evaluation of process performance, the Q-factor [12]. It maximizes the Q-factor, which results in process centering (on target) and improved repeatability (minimum spread). With process specifications and the Q-factor, ABC compensates for off-specification manufacturing processes by making adjustments using "feedforward" for current product and "feedback" for incoming product. This is done via the construction of Q-factor response maps which model the relationships between process variables (inputs) and the Q-factor (output). Adjustment values are chosen to maximize the appropriate Q-factor values. With this capability ABC minimizes scrap and rework, and therefore cost, by correcting poor results from one process step (or process group) in a subsequent step (or group).

ABC provides real-time process-state evaluation which identifies the current process status—*where you are*; the current targets—*where you aimed*; and the optimal targets and optimization path—*where you should be* and *how to get there*. This evaluation, which includes status of sensor-based data as well as non-numeric information about tool performance, raw materials, and other data which affect product quality, is used to continuously optimize the process. Sensor-based data are analyzed using response surface mapping based on regression, neural nets, or rule-based modeling, while "non-numeric analysis" [13] is used to evaluate data of a non-numeric nature. ABC also generates operator-readable control charts which predict process behavior and provide warnings whenever appropriate.

Section 2 defines the new measure of process performance, the Q-factor, used in ABC to evaluate process fidelity and product quality.

Section 3 presents the architecture of ABC. The A-net, B-net, and C-net are described, as well as their uses to perform sensitivity and statistical analyses, to evaluate current behavior and future trends, and to determine optimal process targets.

Section 4 describes ABC's functions and operational flow. The functions fall into five categories: 1) data acquisition and preparation; 2) data evaluation and model testing; 3) improvement of process control and product optimization; 4) detection and treatment of unmeasured but significant process parameters; and 5) process and equipment modeling. These functions are executed in a specified sequence to provide quality analysis and improvement action.

Section 5 treats system design. Topics discussed include ABC databases and their management, ABC logistics and compatibility with manufacturing floor systems, ABC's menu-driven manual interface, and ABC implementation via a network of loosely-coupled workstations which provide load-sharing and high reliability.

The structure and functions of ABC are illustrated using semiconductor manufacturing steps in the critical photolithographic sector. This sector is responsible for the creation of a developed image of the appropriate pattern (circuit) on a preprocessed silicon wafer. The image must be aligned with existing patterns on the wafer. The misalignment error, known as overlay (O/L) error, must be minimized for proper circuit operation.

Section 6 describes a complete application of ABC to the manufacture of a MOS semiconductor capacitor. This example shows how to control and optimize *electrical properties* of capacitors fabricated by a multistep process. It demonstrates ABC's ability to minimize rework by compensating in subsequent steps for out-of-specification outputs of previous steps. The ability to bring processes back into specification is particularly attractive for continuous-flow manufacturing.

2. Q-factor

The Q-factor [12] is a generalization of yield which is used in ABC to measure and improve process performance. It is a sensitive function which describes product quality relative to specifications. A Q-factor exists for each item produced by a process and is integrable over processing time. The average Q-factor for the items produced represents the Q-factor of the process. The Q-factor of a line composed of a series of processes is derived from the Q-factors of constituent processes or is obtained by direct measurements of the final product.

The Q-factor describes the effects of the positive or negative deviations of the item from its target value. The maximum value of the Q-factor, which is normalized to a value of 1, is achieved if the process proceeds exactly as specified and the items produced meet exactly the targeted specifications. Tolerances, which define a window about a target point, are specified for a process in terms of a parameter, ε, which produces an acceptable product.

The properties of the Q-factor are described by a pair of half-Gaussians defined over the positive and negative portions of the target deviations. The half-Gaussians join at the maximum value of the target point. The Q-factor is defined in deviations from target values in units of tolerances, as shown in **Figure 1**.

Formally, let the mean deviation from a process target be m and the process variance be s^2, all given in tolerance units. Then,

$$Q = \frac{1}{\sqrt{1 + 2ks^2}} \exp\left(-\frac{km^2}{1 + 2ks^2}\right),$$

where k is a positive number which measures the sensitivity of the Q-factor to variability and to deviations from a target value. The parameter k is set to achieve any desired sensitivity at the tolerance limits, such as 3σ or 6σ.

If the desired process is specified for a mean that is exactly on target, and a standard deviation with a tolerance limit equals three standard deviations (i.e., $m = 0$ and $s = 1/3$), setting $k = 9/2$ results in a Q-factor of $1/\sqrt{2}$ for that process. For a 6σ process, which is a process in which the tolerance equals six process standard deviations, setting $k = 18$ also results in a Q-factor of $1/\sqrt{2}$ for the process. Thus, the Q-factor can be designed to be independent of the number of standard deviations of the produced items that are desired to fall within the tolerance window. The same Q-factor can be used for a 3σ or 6σ process.

The aforementioned procedure defines Q as a variable with value 1 if all product is on target ($m = s = 0$) and a value of $\varepsilon = e^{-k}$ if all product is at tolerance limits. For a 3σ process, $\varepsilon = 0.011$; for a 6σ process, $\varepsilon = 1.52 \times 10^{-8}$.

For an individual measurement $s = 0$, the Q-factor simplifies to

$$Q = \exp(-km^2),$$

where m is an individual measurement deviation from a target value given in tolerance units.

3. ABC architecture

A general ABC network composed of the A-, B-, and C-nets is shown in **Figure 2**. This hierarchical description of manufacturing enables outputs, or responses of processes at lower levels, to become input process variables at the current level. **Figure 3** shows an ABC network which represents a photolithographic sector in which the semiconductor manufacturing steps start with a prepared substrate and culminate with a developed and tested image of the desired pattern on the substrate. The A-net, which comprises six process steps, is locally controlled by a four-node B-net and is globally optimized by a two-node C-net. Control and optimization operate recursively from level to level. Evaluation of process status and improvement paths is continually performed using the Q-factor as the quality measure.

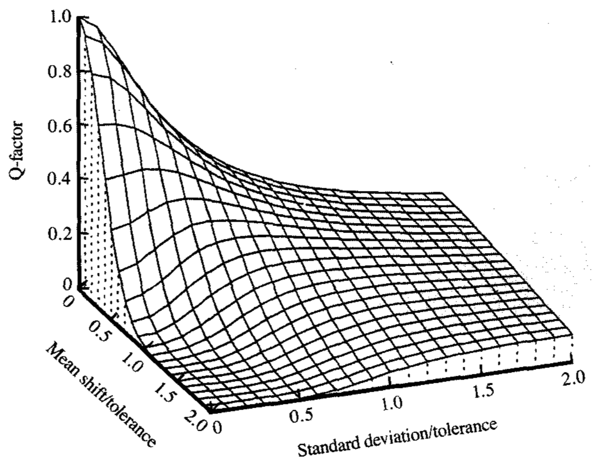

Figure 1

The Q-factor is shown as a function of process "spread" (standard deviation), and process mean shift from target, measured in units of process tolerance.

- *Models*

ABC modeling is a dynamic process. The models are automatically updated and modified via on-line data analysis as new data become available. In addition to changes in parameter values, model modifications include the selection of relevant variables and deletion of unnecessary ones.

The model for each manufacturing process at any level consists of all relevant inputs, outputs, specifications, constraints such as physical limitations, reaction times to adjustment commands, and interconnections which fully describe the process. All relevant process steps, sectors, manufacturing lines, equipment, and partial and complete products are modeled by the A-net. It explicitly defines the

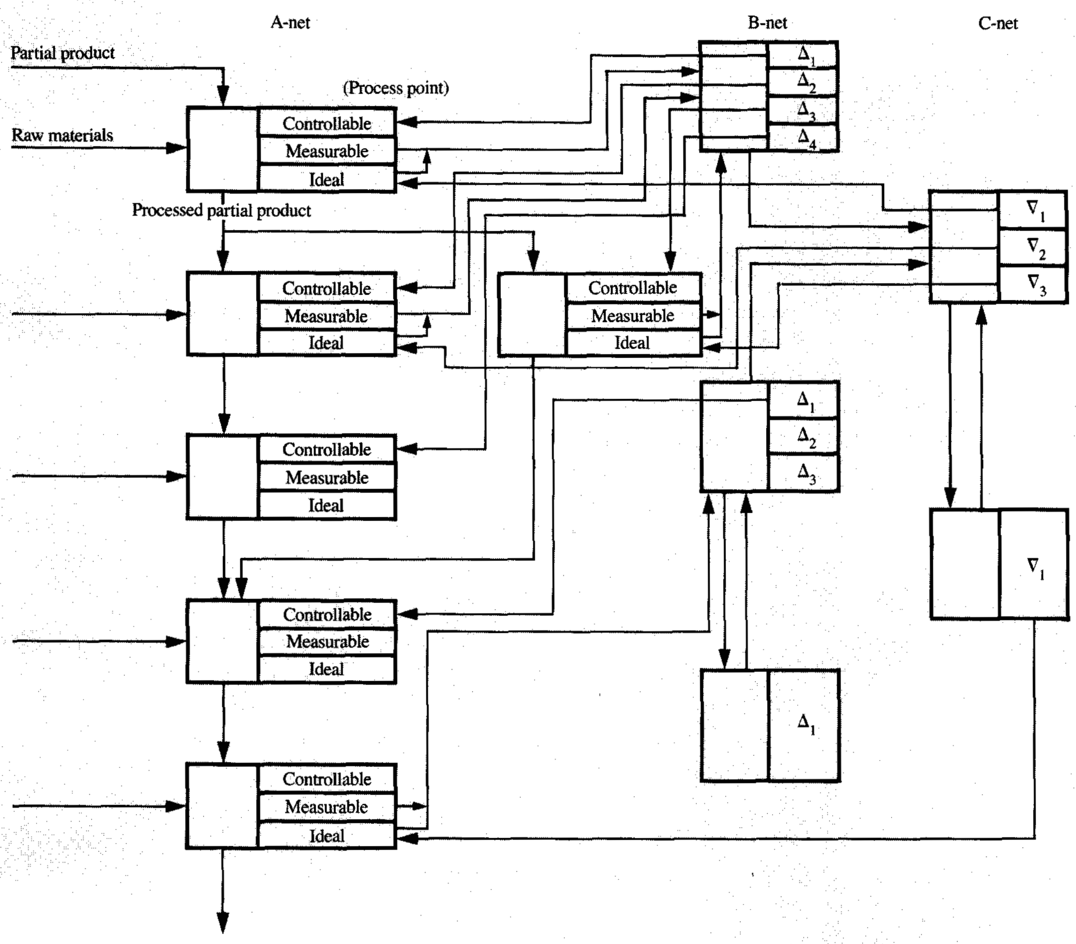

Figure 2

General ABC network structure and interconnections. The A-net models the processes. The B-net gets data from the A-net measurable variables and uses the Δ operators to adjust the A-net controllable variables for local control. The C-net uses the ∇ operators to send optimization adjustments to the A-net ideal (target) variable for global (product) optimization.

manufacturing process to be controlled and optimized in terms of process parameters and target specifications.

Process modeling is extendible and can readily incorporate new or modified processes. All models employ a minimal set of variables and parameters required for process description. To simplify computational complexity, all manufacturing processes, from single steps through manufacturing sectors and lines to complete factories, are treated identically.

Process models are treated stochastically. The models themselves may be analytic, with given functional forms, or empirical, in which case second-order (or higher-order) multivariate polynomial forms [14] are used. Time-evolution models, which may be highly nonlinear, are used in forecasting process behavior and are based on standard time-series analysis techniques, e.g., autoregression [15].

A default quadratic form for empirical data enables the derivation of steps required for process improvement. This is the simplest form which contains surface curvature information and thus enables computation of both the direction and proper step size for process improvement.

Process variables are classified into four basic types: measurable, controllable, ideal, and fundamental. The measurable variables have values that are obtained from product measurements during manufacturing or from tool settings. Controllable variables directly control the process. Ideal variables define nominal manufacturing targets or product specifications. Product and tool parameters can be both a controllable and a measurable variable, depending upon whether this variable is modified to improve the process or used in evaluation of the process. Fundamental variables are derived from basic physical laws that characterize the physical process or product and relate measurable and controllable parameters. These parameters reduce the computation load and generate data from models for which variables are not directly measured.

ABC automatically ranks contributions of variables to process models and uses the smallest set of process variables which explain process behavior within a given confidence level that enables process improvement. These are termed primary variables. All other variables are stored in a secondary variable set for possible future transfer to the primary set if they are needed for the process model. As processes change in time, significant secondary variables are thus automatically transferred by ABC to the primary set, and insignificant primary variables are transferred to the secondary set. Furthermore, if significant variables are not measured, ABC alerts the user to the evaluation of missing variables and evaluates the contributions of newly monitored variables, adding them to the primary set if appropriate.

ABC normally updates process models using "business-as-usual" data from the floor, producing minimum interference with manufacturing operations. At times, the data are insufficient to produce statistically significant results, and more drastic action is required. In such cases, ABC goes into an experimental design mode and requests a series of measurements to enable proper updating of the model.

In many manufacturing processes, data required for real-time analysis are not available when they are needed. In these cases, ABC infers the required data from the latest available models, e.g., equipment models. These models are then updated with actual measurement data when they are available.

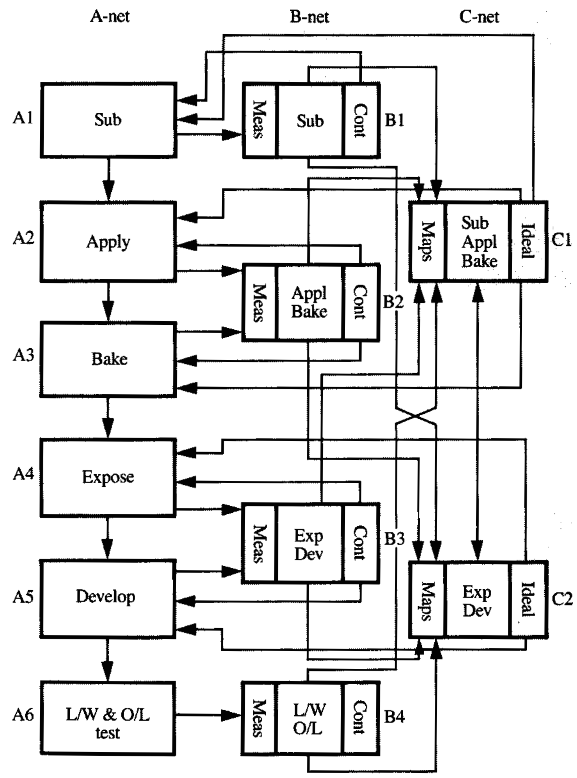

Figure 3

Example of ABC network for the photolithographic sector in semiconductor manufacturing. A substrate (Sub) is coated with photoresist. A circuit-pattern is then exposed through a mask, and the image is developed and tested for linewidth (L/W) of the exposed circuit and overlay (O/L) accuracy of the circuit-pattern placement relative to the underlying structure.

• *Response surface mapping*

Response surface maps [14] model the relationships between process inputs (independent variables) and process outputs (dependent variables) and are used in ABC process control and optimization. It is essential, therefore, that the maps correctly represent the true behavior of all processes.

ABC uses three basic techniques to develop reliable process response maps: regression analysis, neural net analysis, and rule-based modeling. Regression analysis is used when the *form* of the input/output relationship is known, e.g., linear, quadratic. In this case the relationship is explicit and simple, with easily computed confidence levels.

Neural nets are used to develop the relationship in those cases where there is no *a priori* knowledge of the form of the input/output relationship. In such cases, the developed relationship is explicit but need not be simple and is not normally known in closed form. Confidence levels are less reliable. They depend, in part, on the convergence of this iterative technique, which requires a training phase.

Rule-based modeling, using fuzzy inference, is used if the previous two techniques yield unsatisfactory results. This artificial intelligence tool requires initial knowledge input from process experts. It models the input/output relationships as a set of fuzzy rules, and refines the initial rule set using process data. Here, the process response surface is implicit and is captured in the set of rules. The confidence levels can only be attained via comparison with actual results.

A-net		Variable Name	Variable Type	Description
A1	Substrate	X_{11}	M, I	Wafer history
A2	Apply	X_{21}	C, M, I	Time, temperature, speed
		X_{22}	M, I	Resist, parameters
		Y_2	M, I	Thickness
A3	Bake	X_{31}	C, M, I	Time, temperature
		Y_{31}	M, I	Optical parameters
		Y_{32}	M, I	Thickness
A4	Expose	X_{41}	C, M, I	Tool settings
		X_{42}	C, M, I	Dose, focus, mask
		Y_4	M, I	Latent image
A5	Develop	X_{51}	C, M, I	Time, temperature
		X_{52}	M, I	Develop parameters
		Y_5	M, I	Develop image
A6	L/W & O/L test	Y_{61}	M, I	Linewidth
		Y_{62}	M, I	Overlay

Figure 4

Example of A-net in the photolithographic sector and its variables. Input variables (X_{ij}) and output variables (Y_{ij}) can be M (measurable), C (controllable), or I (ideal). Some measurable variables cannot be controlled, but all controllable variables can be measured.

• *A-net: Manufacturing modeling*

The A-net represents the manufacturing application. The nodes denote the processes or process-points. (A process-point is defined as a point in the application model at which measurements are made or process control parameters are changed.) The branches define product manufacturing flow comprising the application, and contain coupling weights which specify the correlation of each process to its predecessor. The A-net of the photolithographic sector in semiconductor manufacturing, together with a description of its variables, is shown in **Figure 4**.

A node may represent a simple process step or a complex process defined in turn by its own A-net. An example for the photolithographic sector is the starting substrate, which undergoes several prior steps required for delivery of a prepared substrate to the photolithographic sector. The product flow can branch to parallel processes, i.e., those that do not have sequential dependencies on one another, and thus can operate in parallel, and also can loop back to itself or to prior processes to represent "product rework."

Although equipment and processes are handled in a similar manner (both have inputs, outputs, and models), they are partitioned into separate models by ABC. This enables the specific effects of the tools and instruments as well as the actual process to be identifiable even though both are combined to form a total process model.

• *B-net: Manufacturing control*

The B-net represents process control. Its primary function is to control the process represented by the A-net to which it is connected. Control in this context means that it evaluates data from an A-net and makes the necessary modifications to the process control parameters in the A-nodes in order to tune the process as closely as possible to its current target values.

Nodes in the B-net denote control operations, and branches specify the flow of control-instructions. Figure 3 illustrates a typical B-net for the photolithographic sector. An example of the functional flow of B-node operators and suboperators, which specifically refers to node B2 of Figure 3, is shown in **Figure 5**.

The B-net is also connected to the optimization C-net (Figures 2 and 3). Information about the status of processes controlled by the B-nodes is sent to those C-nodes. This enables modification of the A-node target values toward global product improvement.

B-nodes contain measurable and ideal variables values which they receive from connected A-nodes. Typically, these consist of job identification tags, job processing start and end times, job size, coded job non-numeric information characteristics (e.g., tools and vendors), a number of dependent and independent variables, and sets of dependent and independent variable values.

Control of the process represented by the A-nodes is accomplished by changing A-node control-variables. To achieve this control, certain classes of operations are performed by the B-nodes via operators and suboperators:

1. Data sets are obtained from the A-nodes, checked for correctness or consistency, and, if necessary, "cleaned."
2. The data are clustered and sent to databases.
3. Model parameters represented by the A-nodes are updated with the "clean" data.
4. The updated model is evaluated for consistency with process behavior and, if necessary, modified by incorporation or deletion of variables.
5. Process behavior measures, such as the Q-factor or yield, are evaluated for the adjusted model.
6. Control-variable improvements are computed and evaluated in terms of cost/performance.

7. Control-variables are adjusted if warranted by cost/performance gain.
8. Process behavior and adjusted variables are displayed.

The functional operation of the B-nodes is described in Section 4.

The primary operators of the B-node (denoted by Δs) in turn can call suboperators (Δ_{ij}s). **Figure 6** shows the functional flow of a typical B-node operator, with specific reference to B2 in the network of Figure 3. The flow of the primary operators and their calls to suboperators are invariant over all applications.

- *C-net: Manufacturing optimization*

The C-net represents process optimization. The primary function of the C-node is to optimize the process represented by the A-nodes by changing A-node local target values. Nodes denote target manufacturing-specification adjustment-operations, and branches specify target-specification flow. The C-net operates on data from the A-net via the B-nodes and adjusts the target values of the A-net nodes to improve overall product quality and production performance (see Figures 2 and 3). From the B-nodes, the C-node receives response surface maps, which are based on empirical model building [14] and model process outputs behavior as functions of input parameters, as well as the current settings of controllable and ideal variables.

Certain classes of operations are performed by the C-node:

1. Maps are obtained from the B-nodes, and a cumulative response surface map is created that represents the totality of the processes.
2. An optimum point of the cumulative map is derived, and associated parameter settings are determined.
3. The significance of the improvement, i.e., its cost/performance, is evaluated.
4. The optimum parameter adjustment path is computed if change is warranted, and the appropriate A-node ideal parameters are modified.

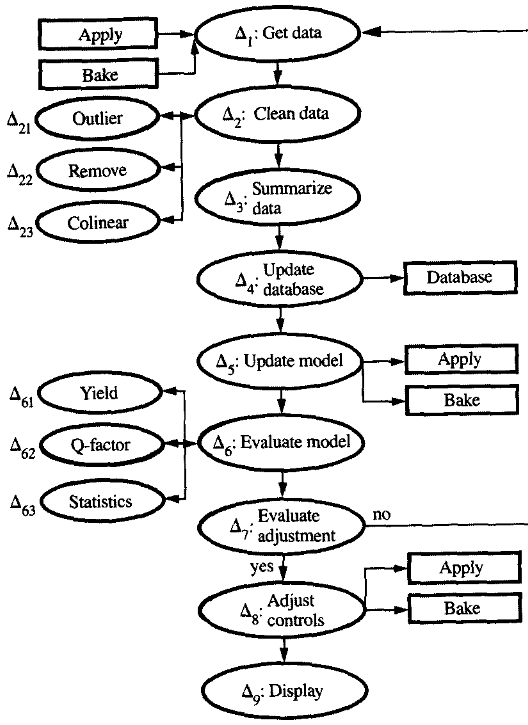

Figure 5

Functional operation of the B2 control node of Figure 2. B2 controls the apply and bake steps of Figure 3 using the Δ_i operators for data analysis, model update and control instructions. The Δ_i operators use Δ_{ij} suboperators to perform specific tasks. For example, Δ_2 (Clean data) calls Δ_{21} (Outlier) to identify suspicious measurements and calls Δ_{23} (Colinear) to identify suspicious variables, and then calls Δ_{22} to remove unwanted data rows or columns.

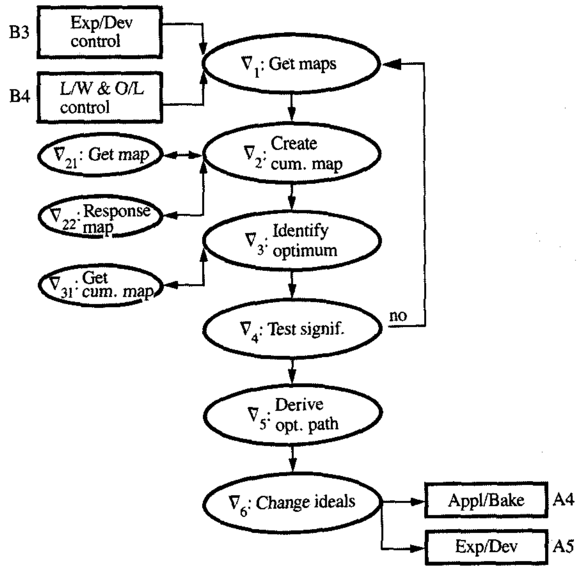

Figure 6

Functional operation of the C2 optimization node of Figure 3. C2 receives information from the B3 and B4 nodes, as well as test results from A6 (via B4). It then uses the ∇_i operators and their ∇_{ij} suboperators to update the ideals (targets) in A4 and A5 so that the quality of the developed image is maximized.

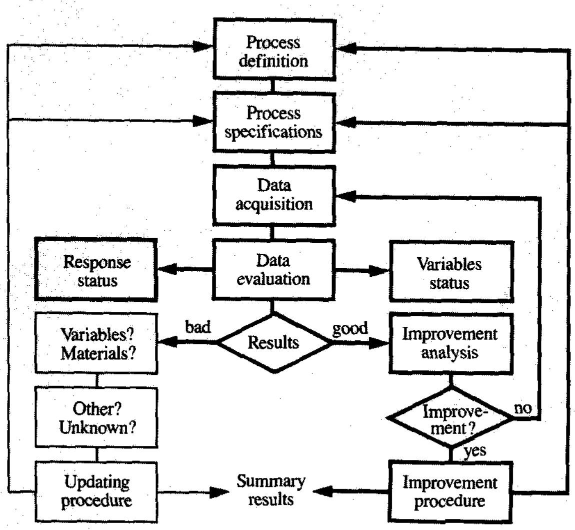

Figure 7

Block diagram of ABC functional flow. Process definition, process specifications, and data acquisition are owned by the A-net. Data evaluation, response, and variable status are performed by the B-net. The B-net is also responsible for secondary analysis which updates process definition with new variables if necessary. Improvement analysis is shared by the B-net (for local control) and the C-net (for global optimization). See text for further details.

Details of the functional operations are presented in Section 3.

The primary operators of the C-node, denoted by ∇s, in turn can call other suboperators which are contained in the generic C-node, along with their associated references, the B-nodes, for obtaining maps and modifying the ideal variables of the A-nodes. Figure 6 shows the functional flow of a typical C-node operator, with specific reference to C2 in the network of Figure 3. The execution flow of the operators and suboperators with the specified input maps generates the modifications of the ideals (targets) in process steps A4 and A5 of Figure 3. The flow of the primary operators and their calls to suboperators, as with the B-node, are invariant over all applications.

• *Optimization flow timeline*

The interaction of the B-net and the C-net to optimize the settings of the A-node ideals is illustrated by an optimization timeline. This timeline employs feedback and feedforward messages to control transmission of B-net instructions (setting A-net control variables) and C-net instructions (setting A-net ideal variables). The timing of these instructions ensures that the appropriate messages are delivered to their destinations as required by product arrival. In general, feedforward instructions are aimed at improving items that are currently produced, while feedback instructions are aimed at improving items produced in the future.

4. ABC functional operation

During each cycle of operation, ABC performs several steps. Each cycle begins with the acquisition of new data and ends with possible control and ideal parameter adjustments and an updated database, which stores the complete process history and status. These steps are now described. A flowchart of the functional operation is given in **Figure 7**.

• *Data acquisition and preparation*

The B-net prepares data for subsequent analysis by ABC. Its function is to ensure data reliability and to code data for ABC use. Observations with incomplete or invalid data are eliminated, outliers are deleted, and colinear variables are removed. Data distribution tests are performed to determine the need for non-normal analysis or transformations toward a normal distribution.

Data are coded in dimensionless form as deviations from the given design specification targets, in units of the given tolerances. During subsequent analysis, variables are directly compared independent of their natural scale of units. For all measured variables, the design target is subtracted from the raw measured variable value and divided by the design tolerance value in order to produce dimensionless, normalized variables for ABC analysis. Major operations performed at this step include summary statistics of raw data; tests and corrective actions for missing and incorrect data; data coding and diagnostics (e.g., for normality, colinearity, and outliers); and data transformations, if needed, in preparation for analysis. At the conclusion of this step, the data used by ABC for process evaluation are reliable.

• *Data evaluation and model test*

This step by the B-net performs the required analysis for evaluation of process behavior. Cluster analysis is used to identify the set of observations that reflect the current process behavior. Process trends, cycles, and forecasts are computed using time-series analysis techniques [15]. Ranking analysis is performed to detect missing primary variables and insignificant model parameters. Process models are developed or updated and their validity tested. For empirical models, data are regressed [16] on second-order multivariate polynomial forms or, if needed, by neural net analysis [17–19] or a fuzzy rule set, to model process behavior.

Figure 8(a) shows an illustrative response surface of an actual photolithographic process involving the exposure-tool alignment parameters, X and Y, as control variables,

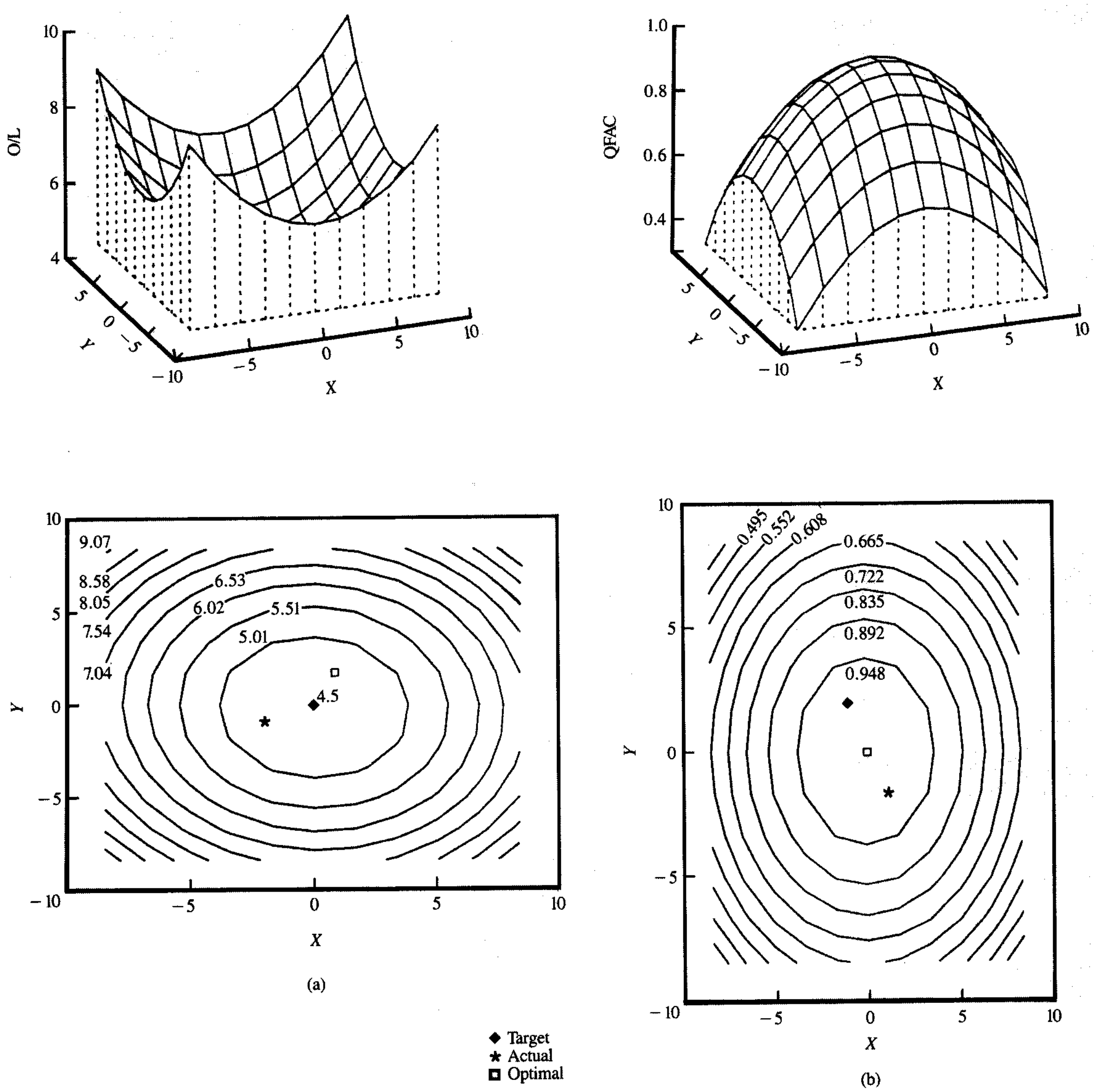

Figure 8

Response surface and contour maps for (a) overlay (O/L) and (b) the Q-factor, in the photolithographic sector, as functions of two input variables (the *X* and the *Y* alignment controls of the exposure tool). These variables are presented as deviations from target values in units of tolerance. Other variables are held at their optimum settings. ABC identifies the optimum input settings (which maximize the Q-factor) as well as process actual inputs and the design target values as shown in the contour maps.

and image pattern overlay error, O/L, as a measurable output variable. It depicts process behavior in the region of interest for the *X* and *Y* alignment parameters. **Figure 8(b)** shows the corresponding map of the Q-factor for the same process, by illustrating the Q-factor mean, target, and optimum for the process. **Table 1** ranks the independent variables with respect to their contribution to overlay behavior. The alignment (X, Y) and rotation (ROT) variables cumulatively control 88.25% of the overlay output and are clearly the significant variables for this process. The other parameters [the isotropic (MAG) and anisotropic (AMAG) magnifications and skew] control only 9% of the process and should be delegated to secondary variables. All six variables explain 97.5% of

Table 1 Photosector variable ranking of the O/L response.

Variable ranking	*Individual contribution (%)*	*Cumulative contribution (%)*
X	42.2	42.2
ROT	27.3	69.5
Y	18.6	88.2
MAG	4.8	93.0
AMAG	3.5	96.6
SKEW	0.8	97.5

overlay behavior, so that all relevant variables for overlay are indeed monitored.

On the basis of analysis, ABC determines the subsequent course of action by the B-net. This may be "improvement analysis" if models are consistent with data, or "secondary analysis" if important parameters are missing. If the analysis is inconclusive, ABC requests additional data and possibly design-of-experiments on all independent process variables which do not exhibit sufficient variability required for model update (Figure 7). At the conclusion of this step, the current model is reliable, and confidence in the recommended B-net subsequent action is high. Major operations performed at this step include classifications of variables via cluster analysis [20]; evaluation of trends, cycles, and forecasting via control charts, followed by variable interrelationships via ranking analysis (i.e., significance of variables, and determination of missing variables); and creation of process response surface maps which identify current process status.

• *Improvement analysis*

If data and models are consistent and reliable, ABC investigates potential improvement. Optimum conditions are derived on the basis of process response surfaces, and their statistical significance is evaluated. The cost of making process changes is compared with the incremental gain in value and net profit of the product. Feedback and feedforward information from processes external to the one being controlled is considered, and subsequent action is determined from the cost-to-profit ratio, assuming that it is statistically significant. Expected process response under nominal, current, and optimum conditions and an optimized path to effect process improvement are computed.

The B-net uses the updated model from the previous step to determine the actual process state (*where you are*) and the nominal process state (*where you aimed*). If the difference is significant, it changes the settings of the controllable variables in the A-net to enable the process to approach its nominal state.

The C-net uses the same model to determine the process optimum state, *where you should be*, and the nominal state *where you aimed*. If the difference is significant, the C-net changes the nominal specifications or the ideals in the A-net, subject to manual override, and alerts manufacturing personnel that improvement is possible via changes in current specifications.

Adjustment of the controllable variables by the B-net and of the ideal variables by the C-net to improve the process compensates not only for perturbations of the current process, but also for the ill effects of other manufacturing processes, thereby minimizing the need for product rework. Feedback for incoming product and feedforward for current product provide the necessary mechanism.

The adjustment of process variables in a multilevel process is performed recursively. The treatment is similar to a composite function which starts from an initial measurable parameter and terminates in controllable parameters at an elementary level. The independent variables of one level are dependent variables at the previous level until elementary controllable variables are adjusted.

At the conclusion of the "improvement analysis" step, the controllable variable settings under B-net control and the ideal variable settings under C-net control are updated as described in the subsection entitled *Process and equipment description update*. Major operations performed in this step include estimation of optimum process conditions via the process response maps and the expected response at optimum; determination of the optimization path (i.e., step size and direction); and update of process targets using feedforward for current product and feedback for future product (i.e., process adjustment for incoming product).

• *Secondary analysis*

If new data do not correlate well with the model as defined by the current process behavior, ABC examines secondary variables for possible improvement of the model. Correlations between secondary variables and process response are evaluated by the B-net, as well as the secondary variables contributions to the multiple correlation coefficient of the model. The set of appropriate secondary variables are automatically transferred into the primary set if the significance threshold is exceeded. Similarly, primary variables which are found to be insignificant are eliminated from the process model, and are transferred to the secondary set for possible future incorporation in the model as primary variables.

At the conclusion of this step, the sets of primary and secondary variables are updated to reflect the latest changes in process models, and appropriate information is sent to manufacturing personnel. Major operations performed at this step include testing of the significance of trial secondary variables to process modeling and inclusion in the model if appropriate; testing secondary non-numeric variables followed by appropriate corrective actions; and interrogation of the history database for similar patterns and resulting corrective actions.

- *Process and equipment description update*

All process modifications produced by ABC during its operations generate an update of process parameters. This may be an updated model due to recent data, model modifications due to "secondary analysis," or adjustment of the controllable variables resulting from "improvement analysis." Initially, the process is defined by the user. Equipment models are treated in a similar fashion, except that they do not have a secondary variables set.

At the conclusion of this step, equipment and process models are updated to reflect the current state of the process. Major operations performed at this step include description/update of all process and equipment variables (primary and secondary), including the response variable; description/update of all process/equipment constraints and process and tool specifications; and all intra- and inter-process correlations for feedback and feedforward adjustments.

- *Non-numeric data evaluation*

The performance of manufacturing lines is directly affected by variables, such as specific tools and vendors, which are non-numeric in nature. These have a strong effect on product quality and must be controlled together with the normal numeric sensor-based data, such as temperature and thickness, to achieve total quality control. A method has been developed [13] to quantitatively evaluate and rank the contributions of these non-numeric attribute categories to process performance such as yield, percent defects, or other measures of process quality.

The technique performs deconvolution (unraveling) of the effects of specific non-numeric variables on the quality measure and automatically identifies the "low performers" responsible for quality degradation. The deconvolution is necessary because the quality of a production run is controlled by a convolution (mixture) of many non-numeric contributers. This is a unique and very important aspect of quality control and must be taken into account in control and optimization of manufacturing.

5. System design considerations

- *Databases*

ABC maintains small, flexible file structures which are summary databases designed for speed and ease of access. One database for each process group, or subset of A-nodes representing a natural grouping of process steps, provides fast response for on-line operation. The databases are updated with on-line data and information available from off-line sources. They contain historical process information which includes a data summary and process-performance measures, specifications, and models, as well as a history of problems. Raw data are stored off-line for future audit, but are normally not referenced by ABC.

Summary information is stored on a cluster or job basis. The file is identified by a job number. It contains the job start and end times for analysis, file search and retrieval purposes. File characteristics, such as job size, number of dependent and independent variables, and coded information pertinent to the job lot, such as tool-operator name, tool-number, name of materials supplier, and source of partial product, are included. The set of Q-factor values for pertinent dependent variables to measure process performance, the yield of the processed product, the set of expected values and variances, and minimum and maximum values for dependent and independent variables also are included.

ABC may have to interact simultaneously with other databases to secure historical and current off-line data. Accessing manufacturing data is particularly important when ABC is used as an off-line tool. Databases may be hierarchical, tree-structured, relational-structured, or of any type. For example, SQL/DB2 (Structured Query Language/Database) is an information retrieval system with a relational database (DB2®) that is of interest for ABC, and whose structure is different from other databases of interest.

ABC interacts with the databases via a database interface module, DBIM. The DBIM translates the request for information and receives data from the appropriate database via one of a set of information retrieval systems (IRSs) which are directly connected to databases of a given type (e.g., tree-structured, relational). The task of a particular IRS is to communicate directly with the database and supply the data to the DBIM, which translates it to a flat-file in a correct format for ABC. New databases of varied types can be rapidly interfaced with ABC with little difficulty. The next section addresses the mode of connecting the DBIM to the IRSs in the context of connecting ABC, in general, to the equipment and systems in a manufacturing environment.

- *Logistics*

ABC is implemented both as a system that is not directly connected to the manufacturing equipment but works through databases, and as an on-line real-time system that directly monitors and controls the manufacturing processes. A workstation or a mainframe host provides a suitable link for ABC as an off-line manufacturing tool. A loosely-coupled network of workstations, such as personal computers in conjunction with dedicated "built-in" toolmicroprocessors, is the preferred system for on-line use. Both implementations accept manual inputs via a menu-driven interface for queries and information. Both also connect to historical product and process summary databases.

Measured data for analysis and parameter value settings are normally entered as files from a database to control and optimize a process. Initial parametric values are set by invocation of data files. ABC analysis results and process recommendations are stored in databases and displayed for manufacturing engineers and equipment operators.

ABC connects to tool equipment and measuring instrumentation as an on-line system and requires a logistic network control system to connect it to manufacturing equipment and databases. ABC is compatible with logistic systems, such as POMS® [21], PlantWorks® [22] and ETSS*. Selection of a logistic system for on-line connection to tools or databases depends on the personal computer operating system. For example, PlantWorks or ETSS is used with OS/2®. The equipment interface is a standard RS232 or IEE488 connector.

Assume that ABC operates in an OS/2 environment under ETSS. ETSS is a PCNET networked-based platform for connecting tools to an application program and for connecting multiple applications, such as a sector controller function. All data transferred between tools are sent via NetBIOS™ commands. All data communications to and from a tool are sent via a SECS protocol using a standard RS232 connector. Communications between an ABC host-based or PC-based system are supported by ETSS. The data needed to drive an ABC application are gathered by ETSS from multiple tools and packaged in a file that can be uploaded to the host or the PC network on which ABC resides.

The control information is packaged in a file by the host or PC network, and is downloaded to the tool network by an application program to change the tool operating parameters in less than ten seconds so as to be suitable for real-time control. POMS, PlantWorks, and any other system are treated in a similar fashion.

- *Menu-driven manual interface*

Manual entry to ABC is through a selection from a menu display. The primary purpose of this menu is to specify the analysis desired from ABC, the data sets, and the processes and equipment. This application is particularly important for an off-line ABC tool. The menu displays and specifications are stored in a database as a historical record of analysis performed by ABC. The database is available for later recall for subsequent process analysis.

As a convenience and safeguard for correct process and/or data set specification, a "mask" is created that controls the particular entries made on a menu. This mask is logically superimposed upon a menu display and is composed of entries with the three values 0, 1, and 2. A value of 0 implies that a change from the previous specification is optional; a 1 indicates that a change must be made; a 2 indicates that no change need be made. The mask is stored in the database, properly tagged, and is recalled for use in future analysis and specification sessions. This artifice is of value for standard production analysis runs such as end-of-the-month reports.

*ETSS is included in an Advanced Operations Module of QuMAP (see [3]).

- *Loosely-coupled workstation network*

The architecture of ABC is implemented in the on-line system mode via a network of loosely-coupled workstations to provide load sharing and high reliability. Dedicated workstations are not required by ABC. Instead, the workload is controlled by a dynamic task management system which evaluates load requirements, processor capacities, and status, and assigns tasks to workstations on a global basis. This approach results in a system that is adaptable to environment changes, flexible enough to permit easy upgradability to changes in manufacturing line designs and process definitions, and capable of incorporating varying numbers of workstations.

In addition to the standard processing tasks, ABC system control employs the concept of a "control task" which is exercised to control the loosely-coupled network. All workstations are candidates for assignment to that task, as well as to the other standard tasks. Upon the execution of the control task, the computer that is assigned acts as a network "master," and the others as "slaves." The other candidates are available in the event of failure of the master; that is, a "dying" master is automatically replaced by a backup master (former slave). The system monitors itself continually.

Task assignments for load sharing are of the two-stage class whereby an initial global assignment is derived first, followed by a dynamic adjustment to reflect recent system load changes.

6. Fabrication of MOS capacitor

To illustrate the structure and functional operation of ABC, an example of MOS capacitor fabrication is presented. This is a multistep process which produces a device that represents a single level in the hierarchy of a semiconductor fabrication line. A one-dimensional model is considered.

Process data were generated by SUPREME II, a Stanford University process simulator, and device data were obtained from VTG, an internal device simulator. Process specifications, where available, were taken from a real semiconductor manufacturing line. Equipment modeling is not included in this example.

The MOS capacitor is designed with a target threshold voltage value of 1 V. The structure and functional operation of an ABC system which monitors, evaluates, controls, and optimizes the capacitor's fabrication process is presented. The actual computed values which are the outputs of the programs used by ABC are shown in the referenced figures and tables and should be followed closely as the text is read.

- *Structure*

Process representation: A-net
The complete MOS capacitor fabrication process consists of six steps: substrate characterization, thermal oxidation, ion implantation, annealing, metallization, and device testing.

Each process step consists of independent variables called inputs, dependent variables or responses called outputs, and models which quantify the relationships between the variables and the responses. The fundamental parameters in all process steps characterize charge distribution which determines device performance. Secondary variables are enumerated but not used in this illustration. The A-net representing MOS capacitor fabrication is shown in **Figure 9**. The process steps are described in the following subsections.

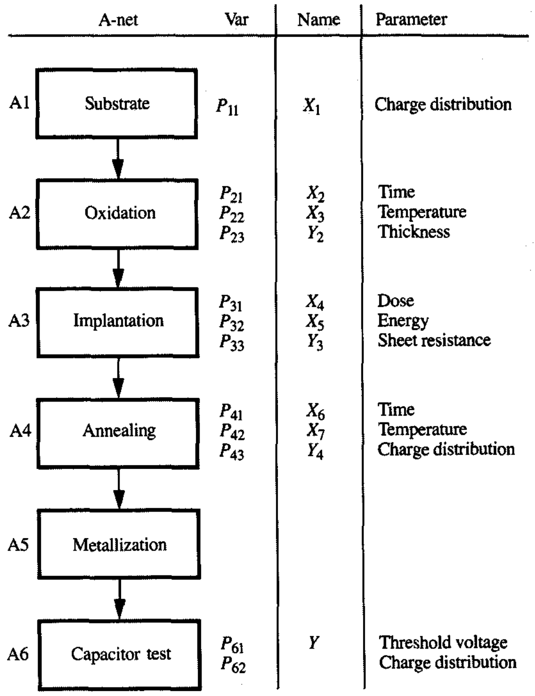

Figure 9

A-net of the MOS capacitor fabrication and its variables. Input variables (X_i) and output variables (Y_i) are described. The Y variables refer to the product's *electrical* performance. All other variables are process parameters.

Substrate (A1) This node represents the substrate characterization process and includes results from all previous manufacturing operations. For MOS capacitor fabrication, the determining characteristic is the charge distribution near the substrate surface. This node is actually represented by an A-net which describes substrate preparation steps such as polishing, cleaning, and doping. For this application, the substrate carries a single primary variable—charge concentration (X_1 in Figure 9), which is considered to be uniform. The specification for this variable is a uniform boron concentration of $(5 \pm 1) \times 10^{16}$ atoms/cc. Secondary variables describe substrate contamination level and flatness.

Oxidation (A2) This node represents the process of substrate oxidation, performed in an oxidation furnace, which "grows" an oxide layer on the substrate. The insulating oxide layer is the "gap medium" of the capacitor, with layer thickness as the determining parameter. For this application, the primary variables are time and temperature of oxidation (X_2, X_3 in Figure 9). The secondary variables are gas composition and partial pressures or flow rate. The primary response is oxide film thickness (Y_2, Figure 9), while secondary responses are dielectric constant and dielectric strength.

An analytic model (the Deal–Grove model) exists for thermal oxidation of silicon. This quadratic model contains two parameters, the linear and parabolic growth coefficients, which are determined experimentally via data fitting.

The specifications for this process step are oxidation time of 45 ± 5 min, oxidation temperature of 800 ± 10°C, and oxide thickness of 24 ± 2 nm.

Ion implantation (A3) This node represents the ion implantation process. It is performed in an E-beam system and is used to control the capacitor's electrical properties via dopant implantation. For this application, the primary variables are ion dose and implant energy (X_4, X_5 in Figure 9). The secondary variables are beam angle and precision parameters. A measurable response variable is the sheet resistance (Y_3 in Figure 9). The fundamental variables are parameters associated with implanted ion concentration. The implanted dopant distribution is modeled by two half-Gaussian profiles with peak and standard deviation parameters.

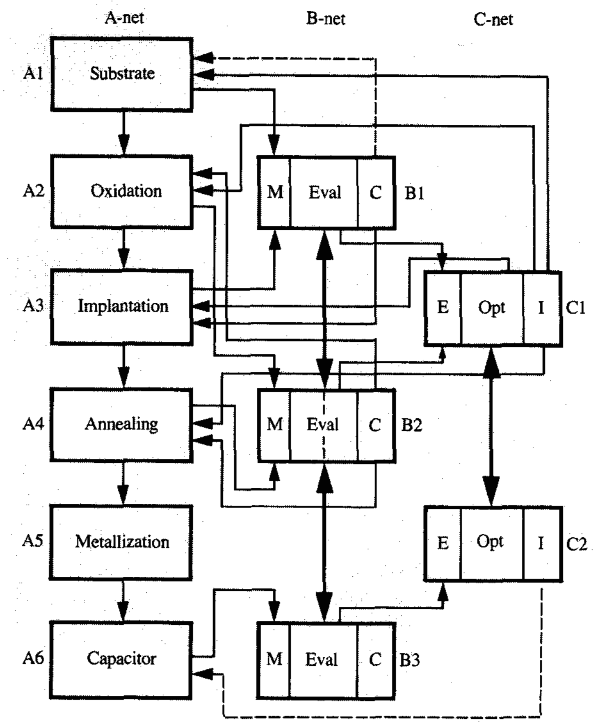

Figure 10

Full ABC network for multistep fabrication of the MOS capacitor. The six A-nodes are controlled by three B-nodes and optimized by two C-nodes. The letters M and C in the B-net refer to the measurable and controllable variables in the A-net. The letter I in the C-net refers to the ideals (targets) in the A-net, while E refers to evaluation information from the B-net.

The specifications for this process step are implantation energy of 30 ± 5 keV, implantation dose of $(1 \pm 0.5) \times 10^{11}$ atoms/cm^2, and sheet resistance of 5240 ± 500 Ω/□.

Annealing (A4) This node represents an annealing process in which substrates are heated in a furnace in an inert ambient. Substrate annealing redistributes the dopant concentration into a stable state. The primary variables are annealing time and temperature (X_6, X_7 in Figure 9). The secondary variables are inert gas composition and pressure. Responses are dopant distribution parameters (Y_4 in Figure 9) which are modeled via one-dimensional diffusion models.

The specifications for this process step are annealing time of 60 ± 5 min and annealing temperature of 1000 ± 10°C. Dopant distribution is not specified for this step.

Metallization (A5) This node represents the process of covering the oxidized structure with a metallic layer. The metallization is performed in an evaporation chamber. For simplicity, this process is assumed to be ideal. Thus, no parameters are required or listed.

Device test (A6) This node represents the process of testing the MOS capacitor. Its primary response is a measurable extrapolated threshold voltage (Y in Figure 9) with a design specification of 1 V. Other measurable parameters such as transconductance and sheet resistance are not used. Fundamental parameters are the acceptor/donor charge distribution parameters which are useful for further processing. Device modeling follows IBM internal simulator VTG, a numerical model.

Process control: B-net

The six process steps corresponding to the six nodes of the A-net are controlled by the B-net, which consists of three interconnected nodes and is shown in **Figure 10.** This logical partition of the B-net reflects the tight coupling between members of each pair of process steps in terms of their effects on the fabrication process. For example, A3 is used to compensate for off-specification doping levels in A1. The partition also reflects similarities in the physics of the processes. A1 and A3 add dopant to the substrate, A2 and A4 are diffusion processes, and A6 tests the product.

The first node (B1) controls the substrate (A1) and the implant process step (A3). The second node (B2) controls the oxidation (A2) and the annealing process step (A4). The third node (B3) receives data from the product (A6) for evaluation. Details of the specific control operations performed by the B-net were described in Section 2.

Process optimization: C-net

The C-net, which optimizes MOS capacitor fabrication, consists of two interconnected nodes, as shown in Figure 10. This logical partition of the C-net reflects the local and global aspects of optimization.

Specifically, C1 is responsible for adjusting the ideal values of individual local processes. This occurs when external considerations, such as design changes, require changes in specifications, or when an optimization procedure, as a result of feedforward data, requires local target changes. It operates individually on ideal variables in A1 through A4 and receives data from B1 and B2.

The C2 node is responsible for global optimization. It receives updated product response maps from B3 and derives optimization paths for the set of ideal parameters which are executed via the C1 node. C2 receives from B3 data such as response surface parameters. The specific optimization operations performed by the C-net are detailed in the following section.

• *Functional operation*

Data

Experimental data for this illustration are generated by a rotatable central composite experimental design using SUPREME II and VTG. There are seven primary variables for this process: one for the substrate, two for oxidation, two for implantation, and two for annealing. The complete design calls for a total of 143 "measurements" which include 128 factorials, 14 axials, and a central point. The data reliability tests of ABC, normally performed at this point, are not required for this example. Data are normalized as part of the experimental design.

Evaluation

Because of the experimental design nature of this illustration, the data clustering for latest process behavior, the time series analysis for process trends, and the control charts normally performed at this point are not appropriate here and are not included.

Ranking and correlation of variables

Following data acquisition, all process variables are ranked in terms of their contributions to the process for the selection of primary and secondary variables. For this illustration, all seven independent variables are initially considered as primary variables, as shown in Figure 9.

Ranking of the variables is performed by B3. The variables are linearly ranked, and variables and their interactions are quadratically ranked. The results of the evaluation show that device threshold voltage is affected by the following factors: oxidation temperature (50%), implantation dose (27%), oxidation time (16%), substrate dopant concentration (4%), and implantation energy (2.5%). The annealing temperature and time parameters contribute less than 0.2% to device performance.

The annealing step has a relatively small effect on the resulting device threshold voltage and is consequently relegated to the set of secondary variables and not used in modeling. The interaction term between the oxidation temperature and the implantation dose, which corresponds to coupling between variables 3 and 5, is found to be as important to process description as the implantation energy, variable 4. Therefore, a linear model for this process is inadequate. A quadratic model describes the process fully and accounts for 99.97% of device performance. The quadratic model is therefore used for control and optimization.

Following ranking of variables, B3 computes all correlations between independent and dependent variables. The results demonstrate that highly ranked variables are also highly correlated with the response. Specifically, the partial correlation coefficient of the oxidation temperature is 0.982, followed by the implantation dose (0.966), the oxidation time (0.946), the substrate concentration (0.826), and the implantation energy (−0.475). The partial correlation coefficients of the annealing time and temperature are −0.068 and −0.173, respectively, and are insignificant. The independent variables are independent and are not correlated with one another. No hidden variables exist in this process.

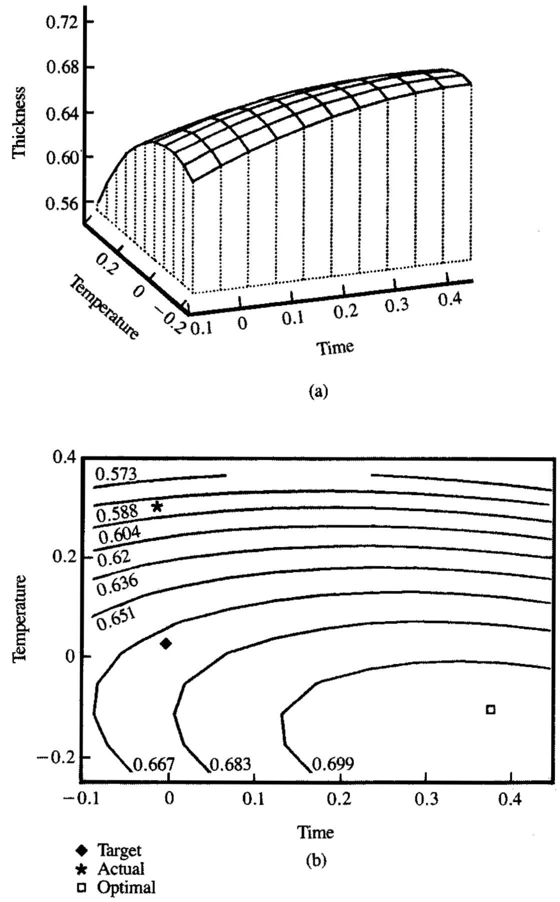

Figure 11

Response (a) surface and (b) contour maps for the oxidation step in MOS capacitor fabrication. Output is oxide thickness as a function of time and temperature of the oxidation furnace. Variables are presented as deviations from target values in units of tolerance. Optimum, actual, and design target values for time and temperature are identified.

Response surface mapping

With reliable data and models, each process step is evaluated and its response surface obtained. In the development of response surfaces, the dependent variables

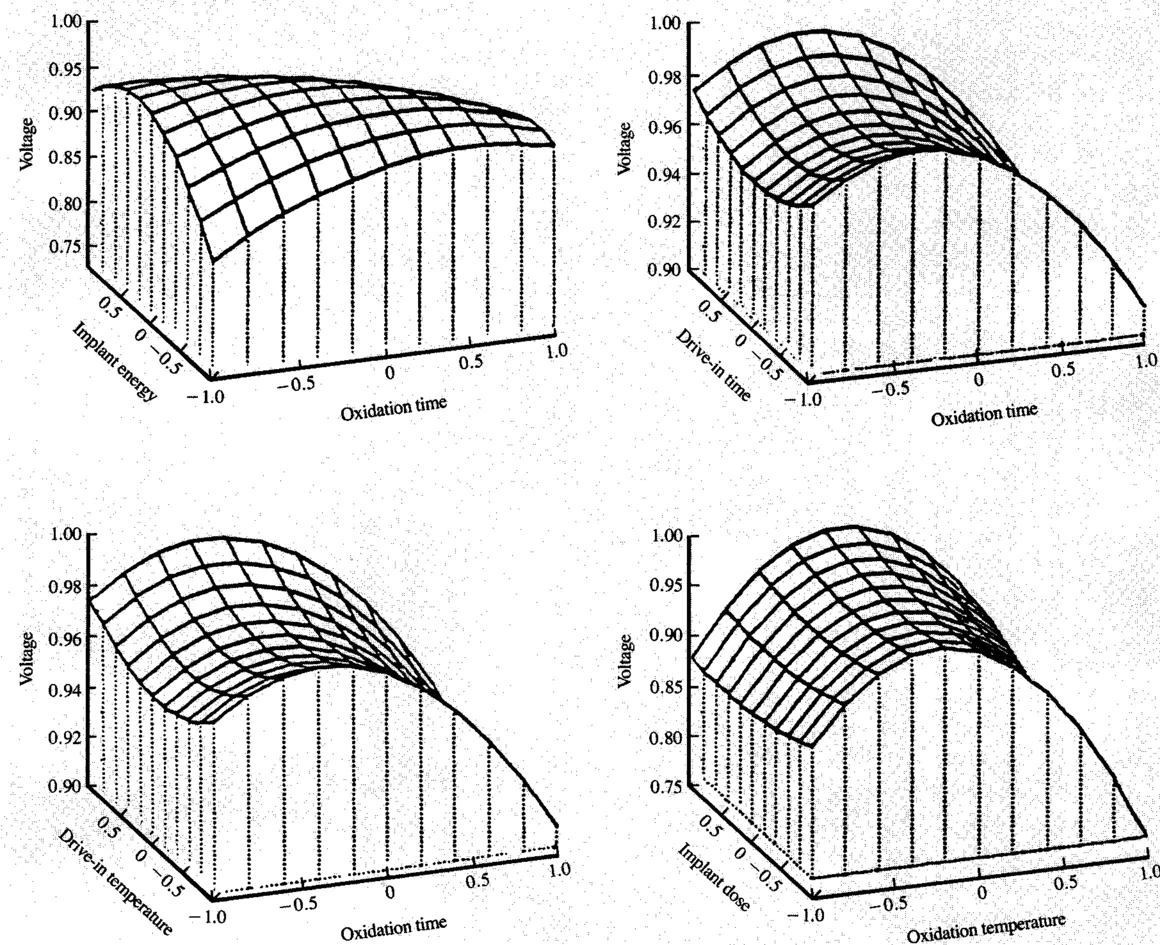

Figure 12

Response surface maps of the MOS capacitor electrical performance (threshold voltage) as functions of several pairs of process variables. Variables are presented as deviations from target values in units of tolerance. All other variables are held at their optimum settings. This figure demonstrates ABC's ability to optimize *electrical* product behavior via adjustments of the controllable variables of *chemical* process steps.

or responses are expressed in terms of their Q-factors. The resulting response surfaces can subsequently be used for optimization. Because of the ranking procedure results, the response surfaces are generated as quadratic functions of their variables.

The response surface of the A2 oxidation step is generated by B2 via regression analysis, and is shown in **Figure 11**. B2 also identifies the coded or normalized coordinates of the nominal state: the time and temperature setpoints (0, 0) corresponding to the decoded values of 45 min and 800°C. The actual state, with current process mean coordinates (0.017, 0.01), corresponds to decoded values of 45.1 min and 800.1°C.

Development by B1 of the response surface of the A3 implant step follows a similar procedure. Here, the coded coordinates of the implant energy and dose at the current process mean are (0.020, 0.022), which correspond to an energy of 30.1 keV and a dose of 1.1×10^{12} atoms/cm^2.

The response surface of the A4 annealing step is not illustrated because of its small impact on device performance. The annealing time and temperature are used in the description of the MOS capacitor performance.

The response surface of the product, A6, is shown in **Figure 12** and describes the performance of the resulting MOS capacitor as a function of all independent variables in capacitor fabrication. The multiple correlation coefficient is

0.9997; the standard error is 0.005. The data fit is at a 95% confidence level. This excellent fit of the model indicates that the relevant capacitor fabrication parameters are included in the model. At the nominal setpoints for all seven independent variables, the device response has a threshold voltage of 1.03 V, which differs from the target of 1.00 V by only 0.03 V.

A product response surface is also generated in terms of the measurable dependent or output variables of individual process steps. For a nominal substrate concentration of 1.0×10^{16} atoms/cc, the Q-factor of the device threshold voltage is derived as a function of oxide thickness Y_2 and implant sheet resistance Y_3, and is shown in **Figure 13**. The coded coordinates of the process nominal state (0, 0) correspond to an oxide thickness of 24.0 nm and an implant sheet resistance of 5240 Ω/□. The process mean state (0.07, 0.02) corresponds to an oxide thickness of 24.3 nm and a sheet resistance of 5263 Ω/□.

Note that the product response map, originally generated in terms of all seven independent variables, is described in terms of three measurable dependent variables: substrate doping level, oxide thickness, and implant sheet resistance. Currently in IBM manufacturing, no measurements are performed after the implant step. It is clear that implant sheet resistance could be measured to improve the fabrication of MOS capacitors.

Control

Following evaluation, the B-net computes a control path from the current state of the process to its nominal state. The nominal state is defined by the current settings or setpoints of the ideal variables. This illustration uses a single-step control path.

B2 computes the positions of the actual and nominal oxide thickness Q-factors for the oxidation step and determines the adjustments required to control the oxidation time and temperature, as shown in Figure 11. Here, the coded adjustments are −0.017 for the time variable and −0.01 for the temperature variable. These new settings are decoded by B2 and reduce the time by 0.085 min and the temperature by 0.1°C. B2 then updates their control settings in A2.

For the implantation step, B1 follows a similar procedure, which results in an implant energy reduction of 0.1 keV and an implant dose reduction of 0.1×10^{12} atoms/cm^2.

Optimization

Optimization consists of making adjustments to the setpoints or ideal variables, employing feedback to improve future product, and using feedforward to bring off-target product back on target. The settings of all ideal parameters in the A-net are performed by C2 via C1. As shown in Figure 13, the optimized state of the MOS

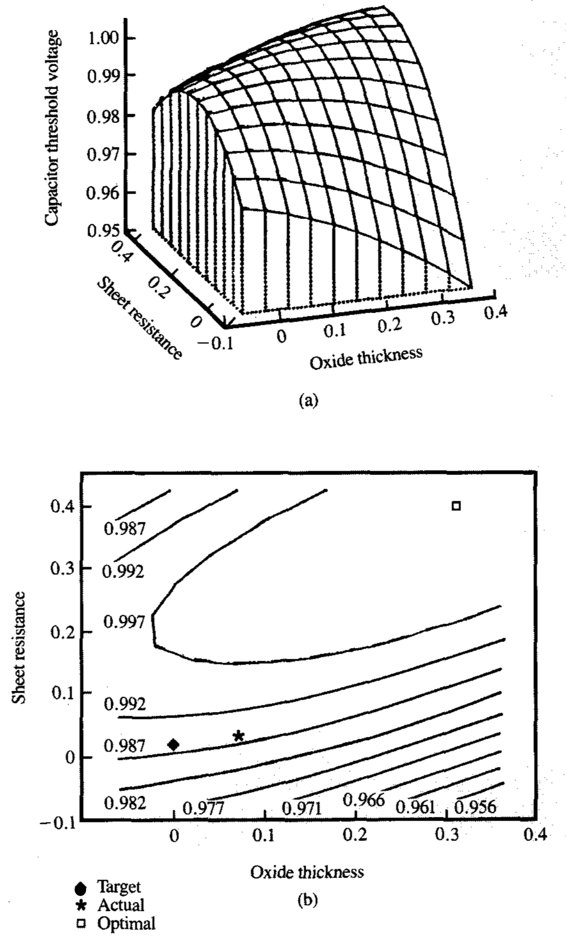

Figure 13

Response (a) surface and (b) contour maps of the MOS capacitor electrical performance (threshold voltage) as functions of oxide thickness and implantation sheet resistance. Variables are presented as deviations from targets in units of tolerance. The optimum input settings (which maximize the Q-factor) as well as process actual inputs and the design target values are identified. Here, the product is optimized by controlling the *outputs* of the oxidation and implantation steps. The C-net need only set the ideals for step output, while the B-net adjusts the inputs to yield the desired outputs.

capacitor occurs at the coded values for oxide thickness of 0.3 and implant sheet resistance of 0.35. Decoding by C2 results in updated values of 25.2 nm for the oxide thickness and 5590 Ω/□ for the sheet resistance. C1 then updates the corresponding ideal values in A2 and A3.

Feedback and feedforward

In the response surface map of the MOS capacitor, the current setpoints for all process steps result in a threshold voltage of 1.03 V. To achieve the required value of 1 V, C2 computes the proper adjustments to the setpoints. These are substrate concentration of 4.9×10^{16} atoms/cc, oxidation time of 44.9 min, oxidation temperature of 800.1°C, implant energy of 29.95 keV, implant dose of 9.15×10^{12} atoms/cm^2, and annealing time and temperature of 59.97 min and 999.8°C, respectively. Update of the setpoints is performed by C2 via C1, which controls the ideal variables in A1 $\cdots$ A4.

ABC also makes feedforward adjustments for an MOS capacitor in which several process steps are off their targets. To achieve a threshold voltage of 1 V with an initial substrate concentration of 4.8×10^{16} atoms/cc, which is lower than the target of 5.0×10^{16} atoms/cc but within the $\pm 0.5 \times 10^{16}$ atoms/cc tolerance, C2 derives an oxidation time of 44.837 min, which is within specifications of 45 ± 2.5 min. These modified values for the oxidation step bring the capacitor back on the target of 1 V without changes in the following process steps. C1 updates A2 ideals.

Given that the oxidation step results in an oxidation time of 46 min and an oxidation temperature of 807°C, which is off the temperature specification of 800 ± 5°C, C2 adjusts the implantation energy to 52.853 keV to bring the capacitor properties back on target, assuming that the remaining variables are on target. C1 updates A3 ideals.

Given that the implantation step results in an implant energy of 30 keV and an implant dose of 5.6×10^{11} atoms/cm^2 on the energy target but off the dose specification of $(1 \pm 0.25) \times 10^{11}$ atoms/cm^2, C2 calculates that an annealing time of 59.772 min is required to bring the capacitor back on target with a nominal annealing temperature of 1000°C. C1 updates the ideals in A4.

The approach adopted here attempts to correct for off-target values immediately in the subsequent operation. A more conservative approach would distribute the corrective action among several subsequent operations so as to minimize correction sizes in individual operations. The risk in the latter approach is uncertainty in subsequent operations, which may result in conditions that are not correctable. A hybrid approach, maximizing immediate corrections within specifications, is probably an optimum strategy.

Note that ABC eliminated unnecessary scrapping or rework of intermediate product which would otherwise have occurred because of the off-specification oxidation temperature and implant dose. This feature is unique and clearly highly desirable.

7. Summary

A methodology, termed ABC, is described that continually models, monitors, controls, and optimizes product quality during a manufacturing process. Complex manufacturing processes in ABC are managed by utilizing hierarchical, dynamic modeling with recursive control and optimization. An overview is presented of the logical structure of ABC, its architecture, characteristics, functional flow, and system design considerations.

ABC is implemented either as a stand-alone tool on a mainframe or personal computer, or as an on-line, real-time system with its own database on a set of loosely-coupled local processors controlled by a system manager. A functional subset of ABC is currently available as an IBM Program Offering under the name *QuMAP—A Better Control.*

ABC introduces several novel concepts to manufacturing control:

- The Q-factor, a highly sensitive measure of a product's deviation from its specifications, is used to center processes at their target values and reduce process variability to improve product quality.
- Modeling automatically identifies the minimum number of significant process variables, ranks their contribution to the process, and detects the existence of unmeasured significant variables. Real-time response surface process modeling is used to identify process targets, actual process operating points, and optimal process setpoints.
- Off-specifications conditions in a process step are corrected in subsequent process steps via feedforward adjustment instructions, to minimize scrap and rework of current products. Feedback modifications are used to adjust processes for future products.
- The effects of all parameter contributions to product quality—sensor-based and non-numeric—are evaluated for production control and manufacturing optimization.

Examples from the semiconductor photolithographic sector line are presented to illustrate the design and functioning of ABC. The operation of ABC on a multistep process consisting of the fabrication of an MOS capacitor illustrates the capability of ABC to minimize rework by correcting out-of-specification product in a subsequent process step.

The current structure of ABC supports extensions into other aspects of manufacturing, such as production planning, scheduling, inventory control, and cost analysis. Using its three nets and analysis techniques, ABC can create or incorporate models for any manufacturing activity. Specifically, ABC can maximize total profit, the true concern of manufacturing. A profit function is introduced for a surface response map, and product quality, manufacturing costs, and production throughput are then controlled and optimized.

Finally, ABC serves as the basis for a discipline in which a manufacturing process is uniformly defined over different sites in a corporation. This requires a detailed examination and systematic design of processes in terms of explicit specifications of all independent and dependent variables, parameters, and tolerances.

Acknowledgments

We appreciate the enthusiastic support of and interest in ABC by B. L. Crowder, M. S. Pittler, R. C. Lange, B. L. Musits, W. D. Grobman, P. S. Hauge, H. K. Fridrich, G. A. Vassiliadis, A. P. Fabrizio, T. M. Walsh, and M. E. Davis. We thank M. R. Wordeman for suggesting the simulation of the MOS capacitor described in Section 6 and for providing the SUPREME and VTG simulators for capacitor modeling. We thank J. Prins for many of the APL statistical routines used in the prototype APL mainframe implementation.

APL2, AIX, PS/2, OS/2, DB2, and PlantWorks are registered trademarks, and QuMAP and NetBIOS are trademarks, of International Business Machines Corporation.

POMS is a registered trademark of Industrial Computing Designs Corporation.

References

1. J. M. Kurtzberg and M. Levanoni, "ABC: A Better Control for Manufacturing," *Research Report RC-16642*, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, January 16, 1991.
2. J. M. Kurtzberg and M. Levanoni, "Computer Integrated Manufacturing," IBM Docket No. Y0990-004, filed in U.S. Patent & Trademark Office, March 9, 1990.
3. Quality Management and Process Optimization Software (QuMAP), available from IBM Corporation, 1000 N.W. 51st St., Boca Raton, FL 33432. Contact R. Bailey (QuMAP product manager).
4. J. M. Juran, F. M. Gryna, and R. S. Bingham, *Quality Control Handbook*, McGraw-Hill Book Co., Inc., New York, 1979.
5. P. K. Mojumder and A. J. Strojwas, "Statistical Control of VLSI Fabrication Process," *Proceedings of the IEEE/CHMT International Electronics Manufacturing Technology Symposium*.
6. E. Sachs and R. Guo, "Process Control System for VLSI Fabrication," *IEEE Trans. Semicond. Manuf.* **2**, 134–144 (1991).
7. J. L. Zayas-Castro and V. J. Mundo-Garcia, "Yield Improvement Through SPC and DOE," *Ann. Quality Congress Trans.* **46**, 444–450 (1992).
8. M. S. Phadke, *Quality Engineering Using Robust Design*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1989.
9. *Statistical Analysis System (SAS)*, SAS Institute Inc., SAS Campus Drive, Cary, NC 27513.
10. *RS1 for the IBM Personal Computer*, BBN Software Products, Cambridge, MA.
11. E. Yashchin and M. Baskin, "CONTRD/CONTRP A System for Design, Plotting and Monitoring of Serial Data" and "Application of CuSum–Shewhart Control Schemes," *Research Reports RC-11880* and *RC-12339*, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, 1986.
12. J. M. Kurtzberg and M. Levanoni, "Method and Apparatus for Quality Measure Driven Process Control," U.S. Patent 5,196,997, March 23, 1993.
13. J. M. Kurtzberg and M. Levanoni, "Evaluation and Ranking of Manufacturing Line Non-Numeric Information," U.S. Utility Patent Application 07/843,048, February 27, 1992.
14. G. E. Box and N. R. Draper, *Empirical Model-Building and Response Surfaces*, Wiley, New York, 1987.
15. G. P. Box, *Time Series Analysis: Forecasting and Control*, Holden-Day Publishing Co., Inc., San Francisco, 1976.
16. N. Draper and H. Smith, *Applied Regression Analysis*, Wiley, New York, 1981.
17. D. E. Rumelhart and J. L. McClelland, *Parallel Distributed Processing: Exploration in the Microstructure of Cognition*, MIT Press, Cambridge, MA, 1986.
18. N. Bhat and T. J. McAvoy, "Use of Neural Nets for Dynamic Modeling and Control of Chemical Systems," *Comput. Chem. Eng.* **14**, No. 4/5, 573–583 (1990).
19. D. A. Sofge and D. A. White, "Neural Network Based Process Optimization and Control," *IEEE Proceedings, 29th Conference on Decision and Control*, Honolulu, 1990.
20. M. R. Anderberg, *Cluster Analysis for Applications*, Academic Press, Inc., New York, 1973.
21. *Process Operations Management System (POMS)*, Industrial Computing Designs Corporation, Weston, VA.
22. "PlantWorks: Application Automation Edition, The IBM CIM Application Enabler," *IBM Personal Systems Developer*, Fall 1991 issue. Available from IBM Corporation, P.O. Box 1328, Boca Raton, FL 33429. Contact S. Sykes (Worldwide Support).

Received April 20, 1992; accepted for publication September 8, 1993

Jerome M. Kurtzberg *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (retired).* Mr. Kurtzberg received the B.S. degree in mathematics from Temple University in 1954 and the M.S. degree in electrical engineering from the University of Pennsylvania in 1961. From 1955 to 1958 he was with Univac, and from 1958 to 1960 with the Auerbach Corporation. In 1960 he joined the staff of the Burroughs Research Center, Paoli, PA, where he was manager of design automation research. From 1964 to 1967 he was a consultant at the Auerbach Corporation and Kettelle Associates, specializing in operations research problems. In 1967 he joined the research staff at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY. His research areas have included manufacturing control and optimization, plant layout methods, computer design automation, signature verification and handwriting recognition, and modeling and design of multiprocessor systems. He is one of the two creators and developers of ABC. Concurrent with his duties at the Watson Research Center, he served in 1968 as a consultant to the Executive Office of the President, Office of Emergency Planning. Mr. Kurtzberg is a member of the Scientific Research Society of America and a senior member of the IEEE. He chaired the IEEE Technical Committee on Design Automation in 1966 and 1967, and in 1969 was a Lecturer for the IEEE Distinguished Visitors program. From 1976 to 1980 he was an associate editor of *Management Science*. He has published, chaired sessions and organized conferences in his research areas. In 1991 he received an IBM Outstanding Innovation Award for a placement system used to design high-performance computer chips. He has received six other IBM research and invention achievement awards. Mr. Kurtzberg is a co-author of *Design Automation of Digital Systems: Theory and Techniques*, Prentice-Hall, 1972, and of *Microelectronics in Large Systems*, Spartan Books, 1965.

Menachem Levanoni *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (LEVANON at YKTVMV).* Dr. Levanoni received the B.Sc. degree in physics from the Hebrew University, Jerusalem, Israel, in 1964. In 1970 he received the Ph.D. degree in physics (hyperfine interactions and nuclear structure) from the California Institute of Technology. He joined IBM Research in 1971 and worked in a number of areas, including optical guided waves, acoustic surface waves, and acousto-optical and magneto-acoustic interactions. In 1973 he joined the Manufacturing Research Laboratory at Yorktown, working in the areas of ink jet printing and laser applications. From 1982 to 1986 he was manager of Process Control and Instrumentation in Manufacturing Research. Dr. Levanoni is the creator and developer of the IBM Offering "ABC" (A Better Control) software system, also known as QuMAP.