



Détection et localisation de défauts des Wafers par des approches statistiques multivariées et calcul des contributions

Baligh Mnassri, Bouchra Ananou, El Mostafa El Adel, Mustapha Ouladsine, Franck Gasnier

► To cite this version:

Baligh Mnassri, Bouchra Ananou, El Mostafa El Adel, Mustapha Ouladsine, Franck Gasnier. Détection et localisation de défauts des Wafers par des approches statistiques multivariées et calcul des contributions. CIFA 2008, Conférence Internationale Francophone d'Automatique, Sep 2008, Bucarest, Roumanie. pp.Article N° 367. <hal-00338686>

HAL Id: hal-00338686

<https://hal.science/hal-00338686v1>

Submitted on 14 Nov 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Détection et localisation de défauts des *Wafers* par des approches statistiques multivariées et calcul des contributions

Baligh MNASSRI¹, Bouchra ANANOU¹, El Mostafa EL ADEL¹, Mustapha OULADSINE¹,
Franck GASNIER²

¹Laboratoire des Sciences de l'Information et des Systèmes – UMR CNRS 6168
Domaine Universitaire de Saint-Jérôme
Avenue Escadrille Normandie-Niemen, 13397 Marseille Cedex 20, France

²STMicroelectronics ZI Rousset
13106 Rousset Cedex, France

{baligh.mnassri, bouchra.ananou, mostafa.eladel, mustapha.ouladsine}@lsis.org,
franck.gasnier@st.com
<http://www.lsis.org>

Résumé— Dans ce papier nous étudions deux approches dédiées à la détection de défauts dans les puces microélectroniques dans le domaine des semi-conducteurs. Ces approches s'appuient fondamentalement sur des aspects multivariés pour réaliser la tâche de diagnostic du système à partir de données volumineuses et hétérogènes issues d'une chaîne de production. Nous proposons ces approches multivariées qui tirent avantage de leurs capacités de traiter des données complexes et qui permettent de mettre en évidence les effets de corrélation entre les variables éventuellement indécélable dans une approche univariée. Ces thématiques portent sur l'aide au diagnostic de systèmes complexes et l'aide à la décision, plus particulièrement sur des aspects de classification et de réduction de dimensionnalité.

Les résultats de ces approches pour la détection ont été validés avec succès sur des données réelles issues d'une phase de production de plusieurs produits de la société STMicroelectronics. Dans l'objectif de localiser les défauts, nous avons considéré l'approche du calcul des contributions relatives aux deux statistiques de détection de l'analyse en composantes principales (ACP). Pour cela, nous avons d'abord considéré ce qui existe dans la littérature. Cependant, le calcul de la contribution à l'erreur quadratique SPE (Squared Prediction Error) exprimée dans le sous-espace résiduel de l'ACP présente un inconvénient. Ce problème a été évité en proposant une nouvelle forme de cette contribution.

Mots-clés— Diagnostic, wafers, classification, localisation, T2 de Hotelling, ACP, SPE.

I. INTRODUCTION

Les progrès permanents réalisés dans le domaine de l'intégration en microélectronique aboutissent à une complexification des circuits et à la réduction des tailles des puces afin de proposer sur le marché des cartes toujours de plus en plus compactes et performantes. Sur les systèmes actuels (circuits comprenant des millions de transistors, cartes électroniques multicouches), les activités de test et de diagnostic prennent également de plus en plus d'importance tout en étant de plus en plus difficiles à réaliser. Le processus de fabrication de puces électroniques regroupées en *wafers* (galettes de silicium rangées latéralement) est

un processus long et complexe nécessitant un contrôle de qualité répondant à des spécifications bien particulières. Ces *wafers*, regroupés en *lots* (environ 25 *wafers* par *lot*), sont vérifiés notamment au bout du processus de fabrication par un contrôle portant sur l'appartenance à un domaine de validité des différentes mesures de contrôle (mesures de résistivité, de conductivité, *etc*). A la fin de la fabrication, chaque wafer est soumis à une série de tests paramétriques qui permettent d'en vérifier le bon fonctionnement, en testant si les paramètres électriques mesurés d'indice "*j*" ne s'écartent d'une valeur cible (*Target*) située entre la valeur minimale (LSL_j) et la valeur maximale (USL_j) situées elles-mêmes entre deux limites de validité (LVL_j) et (UVL_j). Pendant la procédure de test, les wafers passent par trois phases successives :

- La phase qualifiée de "*hold*" consiste à écarter tous les lots dont certains paramètres spécifiques mesurés sur chaque wafer composant ces lots sortent du domaine de spécifications. Ces wafers sont alors re-analysés par les ingénieurs *PT* (*Parametric Test*) pour confirmer le mauvais fonctionnement du/des wafer(s) (tout le lot) ou attribuer l'alerte à un problème de test.
- Une deuxième phase appelée "*full test*" (tests complémentaires) est réalisée en cas d'anomalie confirmée du/des wafer(s) en retirant ces derniers de la ligne de production afin d'y appliquer des tests plus spécifiques (on parle de wafer *scrapé*).
- La troisième phase vise à tester la fonctionnalité de chaque puce à l'aide du test *EWS* (*Electrical Wafer Sort*) pour garantir le bon fonctionnement selon des spécifications données.

Dans l'état actuel du système de surveillance, si un ou plusieurs paramètres sortent de leurs domaines de validité même pour un seul de ces *wafers*, alors le *lot* entier est retiré de la chaîne de production pour une analyse

complémentaire. Le retrait du lot est communément appelé le *hold* du *lot*. Lors de cette analyse complémentaire (seconde série de mesures), les *wafers* définitivement considérés comme défectueux sont retirés et qualifiés de *scrapés*.

L'objectif de ces travaux est de proposer un système d'aide au diagnostic pour *scrapper* dès la première phase de mesure : tous les *wafers* défectueux tout en minimisant conjointement le nombre de *wafers* incorrectement scrapés (réellement bons), le nombre de *scrap*s non détectés et le pourcentage de lots *holdés* injustement. Ces outils de diagnostic rentrent principalement dans le cadre formel de la classification de données dont nous présentons deux approches :

- La statistique T2 de Hotelling [12] comparée à un seuil de détection qui représente une limite de confiance séparant deux classes différentes. Dans notre cas, les *wafers* qui dépassent le seuil de contrôle correspondent à des *wafers* défectueux (*holdés/scrapés*). Les autres qui sont en dessous du seuil correspondent à des *wafers* sains (non *holdés/non scrapés*).
- Les méthodes de projection : étant donné que la dimension de l'espace des mesures est assez importante, l'Analyse en Composantes Principales (ACP) permet d'une part de réduire le volume des données à traiter et d'autre part d'améliorer les performances générales de classification. L'ACP projette les observations originales dans un nouvel espace orthogonal de même dimension que l'espace d'origine [2]. Cependant, la réduction se résume dans l'estimation d'un sous-espace orthogonal (sous-espace des composantes principales) qui estime le mieux les données originales. Le reste d'informations est exprimé dans un sous-espace résiduel ([5], [6]).

Le résultat obtenu par l'ACP est ensuite utilisé pour la localisation des variables en défauts. Nous considérons l'approche qui consiste à calculer la contribution aux deux statistiques de détection à savoir la SPE et le score distance. Cette contribution donne de bons résultats pour le score distance tandis ce que pour la SPE présente une mauvaise estimation des variables en défauts. Notre travail porte en partie sur la modification de celle-ci en présentant une nouvelle forme de la contribution à la SPE. La partie expérimentale permet d'illustrer la comparaison entre les deux méthodes en considérant les données réelles de ST-Microelectronics.

Ce papier est organisé comme suit : la section II positionnera les problèmes rencontrés par la société STMicroelectronics-Rousset dans la détection et la classification de *wafers* défectueux. Elle présentera en outre le principe de détection des deux approches multivariées utilisées. La première est à base du calcul de la statistique T2 de Hotelling, l'autre est par l'ACP. Un exemple d'illustrations des deux méthodes de détection et de classification sera présenté dans la section III. Dans la quatrième section, nous aborderons l'étape de la localisation de défauts en utilisant les deux approches, connues dans la littérature, de calcul de contributions aux deux statistiques de détection de l'ACP (le score distance et la SPE). Nous finissons ce paragraphe par la présentation d'une nouvelle contribution à la SPE. Enfin, nous concluons dans la dernière section.

II. POSITION DES PROBLÈMES ET SOLUTIONS PROPOSÉES

Notre travail porte sur l'analyse des données issues de tests paramétriques pour proposer une aide au diagnostic sur l'état de fonctionnement des puces électroniques afin de minimiser le pourcentage de pièces défectueuses en sortie de la chaîne de production. Nous envisageons, dans le but d'améliorer les gains de productivité, de trouver un moyen permettant de reconnaître le mauvais fonctionnement des puces dès la première phase de tests sans passer par des tests complémentaires.

Le problème rencontré actuellement c'est que le taux de *hold* (nombre de lots retirés pour l'analyse complémentaire sur le nombre total de lots faisant suite aux tests paramétriques) est trop important par rapport au taux de *scrap* (nombre de *wafers* réellement défectueux après analyse *PT* sur le nombre total de *wafers*). Notre objectif est de réduire au maximum le taux de *hold* et éventuellement de le ramener au taux de *scrap*.

Pour résoudre ce problème, nous avons utilisé deux approches supervisées paramétriques des distributions des données basées sur des analyses statistiques multivariées (T2 de Hotelling, méthodes de projection : l'ACP).

A. T2 de Hotelling

L'utilisation d'une carte de contrôle univariée pour chaque variable de processus peut être une solution dans le cas de variables indépendantes mais dans le cas où les variables sont corrélées, ce qui est notre cas, les résultats obtenus ne sont pas satisfaisants. Ceci nous a amené à développer une approche statistique multivariée. Les cartes de contrôle multivariées sont plus adéquates pour le suivi d'un processus complexe. Leur principal avantage, par rapport aux cartes univariées, est la prise en compte de la corrélation entre les variables. C'est la raison pour laquelle on a utilisé une méthode multivariée pour la détection et la classification des *wafers* basée sur la statistique T2 de Hotelling [12].

Harold Hotelling a été le premier à analyser un système de variables corrélées. Il a développé, en 1947, une procédure de contrôle basée sur le concept de "distance statistique" qui a été nommée la statistique T2 de Hotelling. Cette dernière est une mesure qui représente la structure de covariance d'une distribution normale multivariée. Parallèlement, Mahalanobis a développé une distance statistique similaire connue sous le nom de la distance de Mahalanobis [13].

Une base de données est une matrice dont les éléments sont les mesures de la qualité du produit. Elle est de la forme ($X \in \mathbb{R}^{n \times m}$), où n et m représentent respectivement les lignes (observations) et les colonnes (variables) du processus. Cette matrice s'écrit comme suit :

$$X = [x_1 \cdots x_i \cdots x_n]^T \quad \text{avec} \quad x_i = [x_{i1} \cdots x_{ij} \cdots x_{im}]^T$$

Les observations x_i sont des séries des prélèvements de mesures de m variables structurées dans un vecteur ligne et qui sont prises à différentes périodes de temps i .

L'expression de la statistique T2 est donnée par :

$$T_i^2 = n(x_i - \hat{\mu})^T S^{-1}(x_i - \hat{\mu}) \quad (1)$$

où $\hat{\mu} = [\hat{\mu}_1 \dots \hat{\mu}_j \dots \hat{\mu}_m]^T$ est le vecteur des moyennes estimées, dans notre cas on prend $\hat{\mu}_j = (LSL_j + USL_j)/2$ appelée aussi *Target* [14]. S est la matrice de variance-covariance de la base de données X .

Une hypothèse fondamentale dans la construction de la statistique T2 de Hotelling est que le vecteur d'observation x_i doit suivre une distribution normale multidimensionnelle. La distance T2 consiste à déterminer un seuil de détection représentant une limite de contrôle qui sépare deux classes différentes (les wafers sains et les wafers défectueux).

Si on connaît les paramètres de processus, la limite de contrôle supérieure de la carte de contrôle T2 est donnée par la distribution de Fisher comme suit :

$$UCL = \frac{m(n-1)}{(n-m)} F_{(1-\alpha, m, n-m)} \quad (2)$$

avec m et $n-m$ degrés de liberté, $(1-\alpha)$ représente l'intervalle de confiance de la distribution.

B. Analyse en Composantes Principales (ACP)

L'objectif de l'ACP est la réduction de la dimension du problème en créant un ensemble des nouvelles variables, appelées aussi composantes principales ou scores, qui sont des combinaisons linéaires des variables initiales. Cette technique génère une nouvelle base orthogonale dont les axes sont les composantes principales. Ces dernières structurent les données et rendent compte de la variance maximale, en évitant la redondance de l'information [1]. Ainsi les premières composantes captent la majorité de la variance des données initiales. Elle permet d'étudier les relations entre des variables multiples de manière exploratoire, c'est-à-dire lorsque l'exploitation des données n'est guidée par aucune hypothèse préalable. Son intérêt majeur est d'offrir la meilleure visualisation possible des données multivariées, en identifiant les hyperplans dans lesquels la dispersion est maximale, mettant ainsi en évidence avec le maximum de précision les relations de proximité et d'éloignement entre les variables ([2], [3]).

Généralement, les variables d'un produit sont exprimées par des unités de mesures et des échelles différentes. Pour cela, il est préférable d'appliquer une ACP sur une matrice X centrée et réduite (colonnes de moyennes nulles et écarts types unités).

L'espace orthogonal défini par l'ACP est engendré par les vecteurs propres associés aux valeurs propres (λ_a) de la matrice de corrélation Σ de X . Ces valeurs sont mises dans l'ordre décroissant dans une matrice diagonale notée L :

$$L = \text{diag}(\lambda_1 \dots \lambda_a \dots \lambda_m) \quad \text{avec} \quad \lambda_1 \geq \dots \lambda_a \geq \dots \lambda_m$$

Nous définissons :

- P : la matrice orthonormée de projection constituée par les m vecteurs propres p_a associés aux valeurs propres λ_a de la matrice de corrélation Σ de X :
 $P = [p_1 \dots p_a \dots p_m] \in \mathbb{R}^{m \times m}$ avec $PP^T = I_{m \times m}$
- T : la matrice orthogonale qui représente la projection de X dans le nouvel espace ACP (non réduit de dimension m) basé sur les scores t_a (composantes principales) :

$$T = [t_1 \dots t_a \dots t_m] \in \mathbb{R}^{n \times m}$$

Distances	faible SD	grand SD
grande DO	observations orthogonales	mauvaises dispersions
faible DO	observations régulières	bonnes dispersions

TABLE I

LES DIFFÉRENTS TYPES D'OBSERVATIONS BASÉS SUR LES SCORES DISTANCES (SD) ET LES DISTANCES ORTHOGONALES (DO)

Mathématiquement, l'ACP décompose la matrice X comme suit :

$$T = XP \implies X = TP^T = \sum_{a=1}^m t_a p_a^T \quad (3)$$

Puisque l'objectif de l'ACP est de réduire la dimension de l'espace, les A premières composantes principales ($A \ll m$) sont les plus significatives et suffisent pour expliquer la variabilité d'un processus à travers sa base de données X . Par conséquent, l'équation (3) peut se réécrire de la manière suivante :

$$X = T_A P_A^T + E = \sum_{a=1}^A t_a p_a^T + E = \hat{X} + E \quad (4)$$

où \hat{X} et E sont respectivement l'estimée de la matrice X , à travers les A premières composantes principales, identifiant ce qu'on appelle le modèle ACP, et la matrice résiduelle.

La détermination de ce nombre A sera discutée dans la suite de ce papier. Son idéale valeur est choisie de telle sorte qu'il ne reste pas d'informations significatives dans la matrice résiduelle E . Cette dernière doit représenter une erreur aléatoire [4]. En effet, ce choix doit exprimer la majorité des variances de la matrice X qui peut s'exprimer à travers les valeurs propres les plus significatives.

Les matrices estimées de T , P et L établissant le modèle ACP réduit sont respectivement :

$$T_A = [t_1 \dots t_a \dots t_A], \quad P_A = [p_1 \dots p_a \dots p_A] \\ \text{et} \quad L_A = \text{diag}(\lambda_1 \dots \lambda_a \dots \lambda_A)$$

Dans le paragraphe qui suit, nous focalisons notre étude sur le problème de la détection des wafers défectueux par l'ACP. Pour cela, nous analysons un algorithme appelé : l'Analyse en Composantes Principales Classique, en anglais "Classical Principal Components Analysis (CPCA)".

B.1 L'algorithme CPCA

Cet algorithme consiste suite à une ACP de détecter les observations hors contrôle [5]. Après la projection, chaque observation est repérée par son Score Distance (SD) dans le sous-espace des composantes principales et sa Distance Orthogonale (DO) dans le sous-espace résiduel. Suite à ce repérage, une classification en quatre familles d'observations a été réalisée (tableau 1).

Pour comprendre clairement cette classification, nous nous inspirons de [6] et nous considérons un exemple d'une base de données constituée de trois variables ($m = 3$). Ces données sont projetées dans un espace ACP réduit de dimension deux, il s'agit d'un plan ($A = 2$). D'après

la figure 1, la première famille représente les observations régulières constituant un nuage de points regroupés autour d'un centre empirique situé sur le plan des composantes principales. La deuxième famille constituée par les observations 1 et 4, elles sont situées sur le plan mais loin du nuage de points puisqu'elles possèdent un grand SD et une faible DO . Cependant, la troisième famille (observation 5) est totalement en dehors du plan mais sa projection orthogonale à ce plan rejoint les observations régulières, ce qui implique qu'elle possède un faible SD mais une grande DO . Enfin, la dernière famille est représentée par les observations 2 et 3 qui sont caractérisées par une mauvaise dispersion. Elles sont éloignées des observations régulières et situées en dehors du plan car elles ont des grands SD et DO . Par conséquent, toute observation non régulière est considérée hors contrôle.

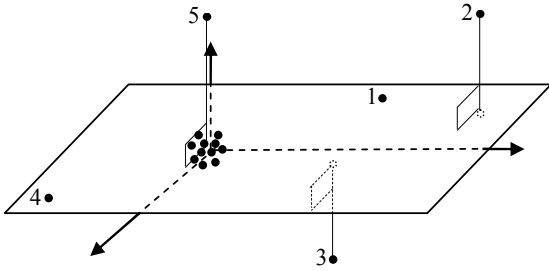


Fig. 1

LES DIFFÉRENTS TYPES DE FAMILLES D'OBSERVATIONS POUR UNE BASE DE DONNÉES DE DIMENSION TROIS PROJETÉE DANS UN ESPACE ACP DE DIMENSION DEUX

B.2 Les statistiques de détection de la CPCA

Après le calcul des deux distances caractéristiques d'une observation projetée dans un espace ACP, l'algorithme CPCA compare chacune de ces deux distances de surveillance à la limite de contrôle de son sous-espace. Dans notre cas, nous notons que le score distance, appelé aussi la statistique D, est calculé par l'approche de Mahalanobis. La deuxième est la distance orthogonale qui est la somme des carrés des erreurs de prédiction connue, dans la littérature, sous le nom de la SPE (squared prediction error) ou la statistique Q ([4], [7], [8]).

Pour une observation i , les deux distances sont exprimées par les équations suivantes ([5], [6]) :

$$SD_i = \sqrt{\sum_{a=1}^A \frac{t_{ia}^2}{s_a^2}} = \sqrt{\sum_{a=1}^A \frac{t_{ia}^2}{\lambda_a}} = \sqrt{t_i^T L_A^{-1} t_i} \quad (5)$$

$$DO_i = \|x_i - P_A t_i^T\| \quad (6)$$

où t_i est la i -ème ligne de la matrice T_A . t_{ia} et s_a^2 sont respectivement la i -ème valeur et la variance du score t_a , avec $\lambda_a = s_a^2$.

Le carré de la distance de Mahalanobis, pour un score qui suit une distribution normale, est approximativement approché par une distribution de Khi-deux ([5], [9]). Par conséquent, la limite de contrôle pour le SD est :

$$LC_{SD} = \sqrt{\chi_{A,1-\alpha}^2} \quad (7)$$

avec A degrés de liberté et $(1 - \alpha) \%$ représente le quantile de la distribution.

Cependant, la détermination du seuil du sous-espace résiduel est plus compliquée car la distribution de la distance orthogonale est inconnue. Par contre, la $g_1 \chi_{g_2}^2$ donne une bonne approximation des distributions inconnues pour la SPE (DO^2). Hubert et al. ([6], [10]) prouvent que la $DO^{2/3}$ suit approximativement une distribution normale de moyenne et de variance exprimées respectivement comme suit :

$$\mu = \left(1 - \frac{2}{9g_2}\right) (g_1 g_2)^{1/3} \quad \text{et} \quad \sigma^2 = \frac{2g_1^{2/3}}{9g_2^{1/3}}$$

Par conséquent, la limite de contrôle du sous-espace résiduel est :

$$LC_{DO} = (\hat{\mu} + \hat{\sigma} z_{1-\alpha})^{3/2} \quad (8)$$

avec $z_{1-\alpha} = \Phi^{-1}(1 - \alpha)$ en tant que $(1 - \alpha) \%$ quantile de la distribution gaussienne Φ . $\hat{\mu}$ et $\hat{\sigma}$ sont respectivement la moyenne et l'écart-type estimés de la $DO^{2/3}$.

III. RÉSULTATS

A. Détection de défauts par la statistique T2

En calculant la statistique T2 de Hotelling des données dont nous disposons (une base de données de grande dimension), nous obtenons les résultats illustrés par la figure 2. La distance T2 de Hotelling arrive à classer un bon pourcentage des wafers Non Holdés/Non Scrapés (wafers sains) en dessous du seuil de détection et les wafers holdés scrapés (wafers défectueux) au dessus du seuil.

L'approche de Hotelling identifie les wafers Non Holdés-Non Scrapés et Holdés-Non Scrapés respectivement avec des taux de 98% et 97% comme des wafers sains et avec un taux de classification global de 97%.

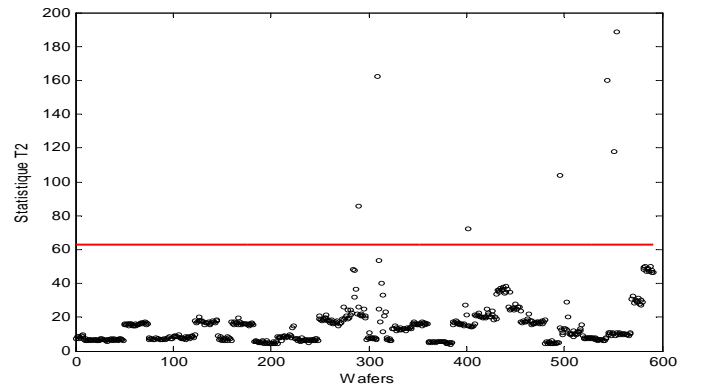


Fig. 2

CLASSIFICATION DES WAFERS SCRAPÉS PAR RAPPORT AUX NON SCRAPÉS À L'AIDE DE LA STATISTIQUE T2 DE HOTELLING

B. Détection de défauts par l'algorithme CPCA

L'étape primordiale dans l'ACP est le choix du nombre des composantes principales (A) à retenir. Puisque ce nombre est directement lié au nombre des valeurs propres exprimant la majorité de la variabilité de la matrice X , il

est donné par l'équation suivante [11] :

$$\sum_{a=1}^A \lambda_a / \sum_{a=1}^m \lambda_a \approx 80\% \quad (9)$$

La même équation est également proposée avec une autre égalité [6] :

$$\sum_{a=1}^A \lambda_a / \sum_{a=1}^m \lambda_a \approx 90\% \quad \text{ou} \quad (\lambda_A / \lambda_1 \geq 10^{-3}) \quad (10)$$

A partir des équations (9) et (10), nous pouvons retenir un nombre de composantes principales donnant une valeur comprise entre 80% et 90% de l'inertie des données initiales de X .

Nous avons réalisé une ACP sur la même base de données utilisée dans la première approche pour tester l'efficacité de cet algorithme. Selon les critères de choix (9) et (10), nous avons retenu 18 (parmi 70) composantes principales ($A = 18$). En outre, avec un quantile de 99% pour chacune des deux limites de contrôle (7) et (8) utilisées pour la détection de défauts, l'algorithme CPCA élabore le graphe illustré par la figure 3. Dans cette dernière, chaque observation projetée est exprimée par ses deux distances données par les équations (5) et (6). Nous constatons un nuage de points en dessous des limites de contrôle de chaque sous-espace et qui représente la famille des observations régulières (observations sous contrôle qui représentent les wafers Non Holdés/Non Scrapés) avec un taux de 95.9%. Les observations restantes sont dispersées aléatoirement d'une manière plus au moins lointaine de la majorité de la population. Elles sont considérées hors contrôle par l'algorithme CPCA et représentent 4.1% de la totalité des données. Nous avons remarqué que les wafers scrapés *PT* appartiennent à la famille d'observations caractérisées par des mauvaises dispersions. Ces wafers défectueux sont les points les plus loin du centre empirique (caractérisés par des grands *SD* et *DO*) et qui sont référenciés par les numéros (2754, 2755, 2717 et 2769) dans la figure 3.

Nous avons validé les deux approches sur des données très complexes issues d'une phase de fabrication de ST-Microelectronics. Pour cette validation, nous n'avons eu aucune connaissance a priori sur les étiquettes (Scrapés et Non Scrapés). Les résultats déclarés par notre système de diagnostic ont été validés par les experts *PT*. Le point commun entre ces deux approches multivariées est la prise en compte de la corrélation entre les variables. Cependant, l'hypothèse sur laquelle est fondée la statistique T2 est que le vecteur d'observation doit suivre une distribution normale multidimensionnelle, ce qui n'est pas toujours possible. Par contre, l'ACP n'impose aucune hypothèse préalable, ce qui rend cet outil plus convenable. Ceci justifie la non détection de quelques observations réellement en défauts par la T2 de Hotelling et qui sont détectables par l'ACP. Cela n'implique pas que cette dernière est parfaite puisqu'elle présente un taux de fausses alarmes non négligeable.

IV. LOCALISATION DE DÉFAUTS SUITE À L'ACP

Une fois qu'un défaut est détecté, il est nécessaire d'identifier les variables qui sont en cause (localisation de défaut).

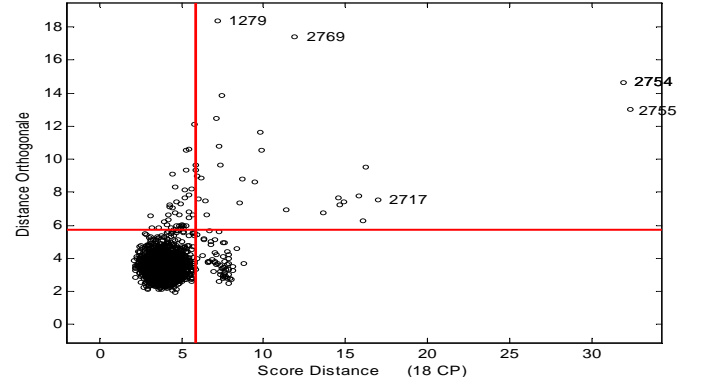


Fig. 3
Détection des wafers défectueux par ACP

Pour réaliser cette tâche, nous utilisons les approches qui consistent à calculer les contributions des variables aux statistiques de détection ([8], [9], [15], [16]).

Avec l'ACP, deux statistiques (le score distance et la SPE) sont utilisés pour la surveillance. Par conséquent, chaque variable aura deux contributions.

A. La contribution au SD^2

D'après l'équation (5), nous remarquons que le score distance est la somme des carrés des composantes principales normalisées. Pour une observation i , chaque composante principale (score) s'exprime comme suit :

$$t_a = p_a^T x = \sum_{j=1}^m p_{aj} x_j \quad (11)$$

où p_{aj} est le j -ème élément du vecteur propre p_a correspondant à la valeur propre λ_a .

De l'équation (11), nous déduisons que la variable x_j contribue dans le calcul du score t_a avec :

$$c_j = p_{aj} x_j \quad (12)$$

Ceci implique que la contribution d'une variable x_j à une composante principale normalisée $(t_a/s_a)^2$ est :

$$cont_{aj} = \frac{t_a}{s_a^2} p_{aj} x_j \approx \frac{t_a}{\lambda_a} p_{aj} x_j \quad (13)$$

Finalement, nous déterminons la contribution totale de la variable x_j dans le calcul de toutes les composantes principales normalisées, c'est-à-dire sa contribution au calcul du SD^2 :

$$Cont_j = \sum_{a=1}^A cont_{aj} \quad (14)$$

Kourti et al. ([9], [15]) proposent de sélectionner les scores normalisés ayant subi une variation significative. En outre, la contribution totale de la variable x_j sera calculée sur les q (parmi les A) composantes principales normalisées les plus élevées. Elle est exprimée comme suit :

$$Cont_j = \sum_{a=1}^q cont_{aj} \quad (15)$$

avec les conditions suivantes : la contribution au score normalisé, donnée par l'équation (13), est mise à zéro si son signe est négatif (son signe est opposé au signe du score t_a). Les variables ayant la plus forte contribution totale sont en défauts.

B. La contribution classique à la SPE

De l'équation (6), nous déduisons que la DO^2 est égale à la somme des carrés des résidus qui représente l'erreur quadratique SPE comme suit :

$$SPE = DO^2 = \sum_{j=1}^m e_j^2 = \sum_{j=1}^m (x_j - \hat{x}_j)^2 \quad (16)$$

Cette approche propose que la contribution d'une variable donnée à la SPE s'exprime par le carré de son résidu ([15], [16]) :

$$CC_j^{SPE} = (x_j - \hat{x}_j)^2 = \left(x_j - \sum_{a=1}^A p_{aj} t_a \right)^2 \quad (17)$$

Les variables qui contribuent d'une façon importante doivent être étudiées.

C. Proposition d'une nouvelle contribution à la SPE

Une observation projetée dans un espace ACP réduit est détectée hors contrôle à travers sa distance orthogonale et son score distance. Par conséquent, une variable en cause sera jugée par ses deux contributions.

En réalité, la contribution d'une variable à la SPE exprimée par l'équation (17) ne représente que son résidu. Cependant, pour déterminer la contribution d'une variable dans le calcul d'une distance quelconque, nous devons extraire la valeur par laquelle elle participe dans ce calcul.

D'après l'équation (17), nous remarquons que le résidu de la variable x_j est exprimé en fonction de sa mesure x_j et de son estimée \hat{x}_j . Cette dernière est en fonction du score t_a qui est lui-même, d'après l'équation (11), en fonction de x_j . Ceci implique que la variable x_j contribue même dans les résidus des autres $m-1$ variables. Pour calculer la contribution d'une variable x_j , il est nécessaire d'exprimer son résidu en fonction de sa mesure x_j uniquement.

Après avoir développé l'équation (16), nous définissons un terme que nous appellerons la contribution relative (R_{uv}) d'une variable u dans le calcul d'un résidu v , notée comme suit :

$$R_{uv} = -x_v e_u \sum_{a=1}^A p_{au} p_{av} = -x_v e_u c_{uv} \quad (18)$$

avec c_{uv} est la $u-ième$ composante de la $v-ième$ colonne de la matrice $C = P_A P_A^T$. Nous pouvons montrer que pour une observation i :

- La contribution de la variable x_j dans le calcul du carré de son résidu e_j^2 est :

$$C_j^{e_j^2} = x_j e_j - x_j e_j c_{jj} = x_j e_j + R_{jj} \quad (19)$$

- La contribution de la variable x_j dans le calcul du carré d'un résidu d'une variable x_r avec ($r \neq j$) est :

$$C_j^{e_r^2} = -x_j e_r c_{rj} = R_{rj} \quad (20)$$

Nous concluons que la contribution totale de la variable x_j à la SPE est la somme de m termes : sa contribution au carré de son propre résidu donnée par l'équation (19), et aux carrés des autres $m-1$ résidus déduites de l'équation (20). Nous exprimons cette contribution totale comme suit :

$$C_j^{SPE} = C_j^{e_j^2} + \sum_{r=1 \neq j}^m C_j^{e_r^2} = x_j e_j + \left(R_{jj} + \sum_{r=1 \neq j}^m R_{rj} \right) \quad (21)$$

Avec le même principe que Kourti et *al.* ([9], [15]) dans leur contribution au SD^2 , nous ne considérons dans notre approche que les contributions positives. Pour cela, nous proposons d'annuler les contributions négatives données par les équations (19) et (20) car elles seront opposées respectivement aux signes de leurs résidus e_j^2 et e_r^2 . Par conséquent, nous proposons les deux contraintes suivantes lors du calcul de la contribution à la SPE (équation 21) :

$$\begin{cases} \text{pour } r = j & \text{si } R_{jj} < -x_j e_j & \text{alors } R_{jj} = -x_j e_j \\ \text{pour } r \neq j & \text{si } R_{rj} < 0 & \text{alors } R_{rj} = 0 \end{cases}$$

Par cette méthode, nous décidons que les variables ayant une forte contribution sont supposées en défauts et doivent être étudiées.

D. Interprétations des résultats

Dans ce paragraphe, nous allons interpréter les résultats de la localisation de défauts suite au calcul des contributions en considérant deux observations quelconques détectées par l'algorithme CPCA. Nous prenons les observations référencées par les numéros 2717 et 2755 dans la figure 3. Pour calculer la contribution au SD^2 , il faut d'abord sélectionner les composantes principales normalisées les plus élevées. Les figures (4.a) et (4.b) affichent respectivement les scores normalisés des deux observations choisies. D'après ces figures, nous pouvons retenir les composantes suivantes : (10, 12, 13, 14 et 15) pour l'observation 2717 et (7, 9, 11, 12, 13 et 17) pour l'observation (2755).

Après le calcul des contributions avec les trois approches (la contribution au SD^2 , la contribution classique à la SPE et la nouvelle contribution à la SPE), nous remarquons que pour le cas de l'observation 2717, les variables supposées en défauts selon la contribution au SD^2 sont : x_6 , x_7 , x_{19} , x_{25} , x_{27} , x_{41} , x_{56} et x_{68} (figure 5). Nous avons constaté, en vérifiant les mesures de ces variables, qu'elles sont réellement toutes en défauts. Cependant d'après la figure 6, la contribution classique à la SPE estime beaucoup plus de variables défectueuses. Outre les variables mentionnées ci-dessus, nous rajoutons les suivantes : x_1 , x_{18} , x_{20} , x_{29} , x_{36} , x_{57} , x_{59} , x_{66} et x_{70} . Nous avons examiné les mesures de ces variables surtout pour celles qui contribuent fortement avec cette approche, exemple : x_{20} et x_{59} . Ces deux variables sont tracées, en fonction des observations en appliquant un zoom autour de l'observation étudiée (ici 2717), dans la figure 8. D'après cette dernière, nous

constatons que ces deux variables déclarées défectueuses par l'approche classique de la contribution à la SPE ne présentent aucun défaut lors de l'observation considérée. Une constatation similaire a été faite pour les autres variables soupçonnées par cette méthode. D'après la figure 7, nous remarquons que notre nouvelle approche pour le calcul de la contribution à la SPE donne des résultats différents à ceux obtenus par celle de la contribution classique. Ces résultats sont pratiquement les mêmes que ceux donnés par la contribution au SD^2 . Pour confirmer ces résultats, nous avons choisi de tracer les mesures de deux variables pour illustrer leurs défauts (exemple : x_7 et x_{27} voir figure 9).

Avec ces deux méthodes (la contribution au SD^2 et la nouvelle contribution à la SPE), nous constatons que les deux sous-espaces générés par l'ACP (celui des composantes principales et résiduel) sont bien exploités afin de dégager une information plus complète et utile pour la localisation. En outre, nous notons que notre nouvelle approche présente un avantage dans la précision du calcul de la contribution par rapport à celle de la contribution au SD^2 , car elle ne nécessite pas la sélection des composantes principales normalisées les plus élevées.

Pour le cas de l'observation 2755, la contribution au SD^2 (figure 10) et la nouvelle contribution à la SPE (figure 12) supposent les mêmes variables en défauts (x_{18} , x_{19} , x_{28} et x_{29}) avec une légère différence au niveau des valeurs de la contribution des deux dernières variables pour la nouvelle contribution à la SPE. Cependant, nous remarquons que la contribution classique à la SPE (figure 11) ne donne pas des résultats cohérents aux réalités physiques des variables. Car selon la figure 11, les variables (x_{28} , x_{33} , x_{36} , x_{68} et x_{70}) possèdent des contributions importantes alors qu'elles ne sont pas réellement en défauts.

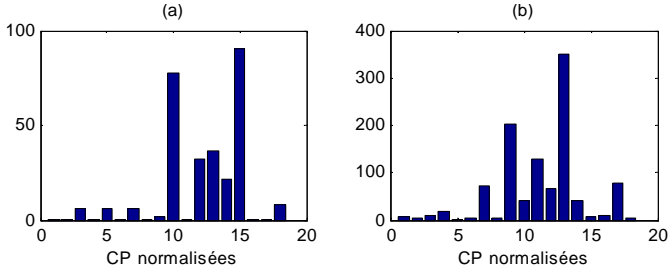


Fig. 4

LES CP NORMALISÉES POUR LES OBSERVATIONS : (A) 2717 ET (B) 2755

V. CONCLUSION

Nous avons présenté une méthodologie de détection de défauts sur des puces microélectroniques à partir de tests paramétriques. Deux approches statistiques ont été développées et testées séparément (la T2 de Hotelling et l'ACP). Ces deux méthodes prennent en compte les corrélations entre les variables et permettent une classification des wafers. En outre, elles présentent une haute sensibilité pour la détection de défauts.

Après avoir testé et validé ces deux approches sur des données réelles sans la connaissance des décisions prises par les ingénieurs PT "Hold et Scrap", nous avons abouti

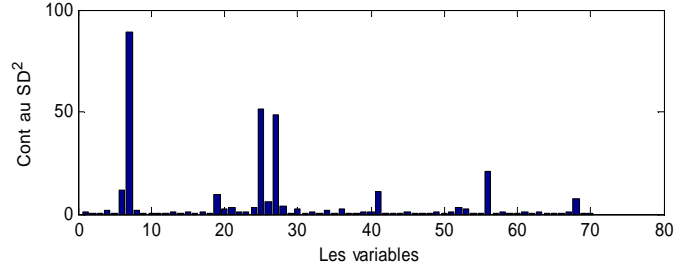


Fig. 5

LA CONTRIBUTION AU SD^2 POUR L'OBSERVATION 2717

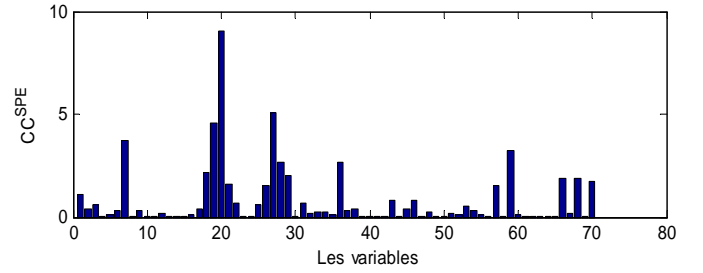


Fig. 6

LA CONTRIBUTION CLASSIQUE À LA SPE POUR L'OBSERVATION 2717

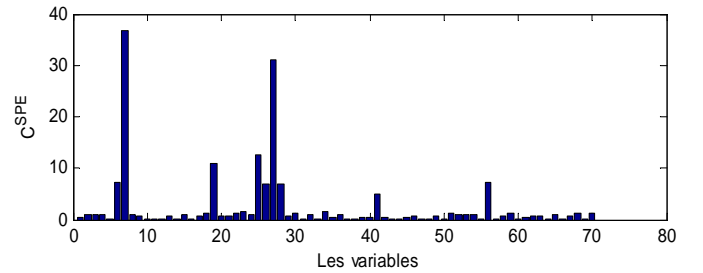


Fig. 7

LA NOUVELLE CONTRIBUTION À LA SPE POUR L'OBSERVATION 2717

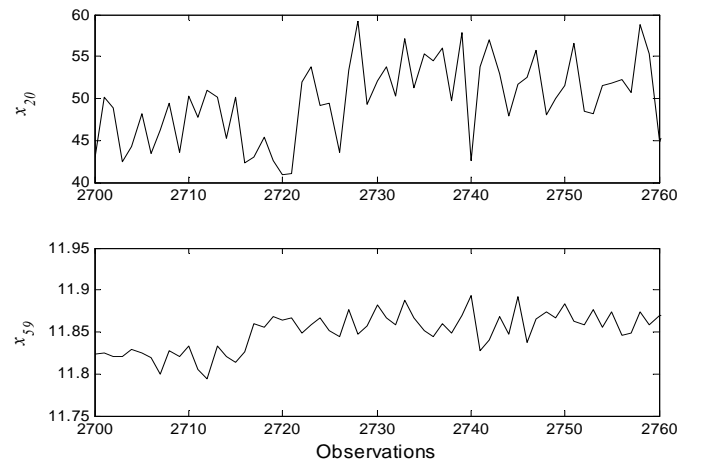


Fig. 8

LES MESURES DES VARIABLES x_{20} ET x_{59}

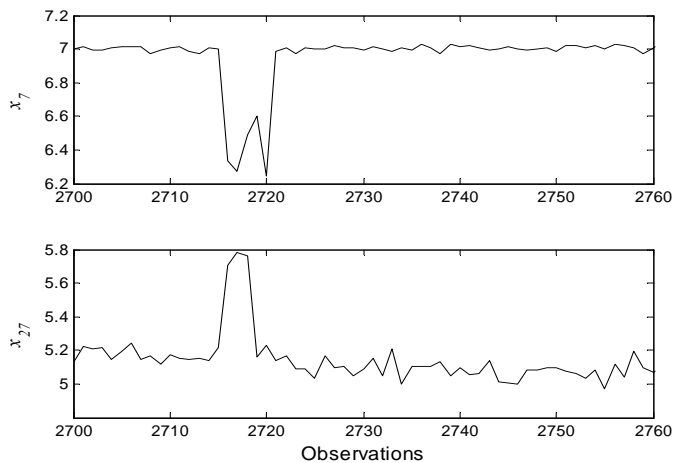


Fig. 9
LES MESURES DES VARIABLES x_7 ET x_{27}

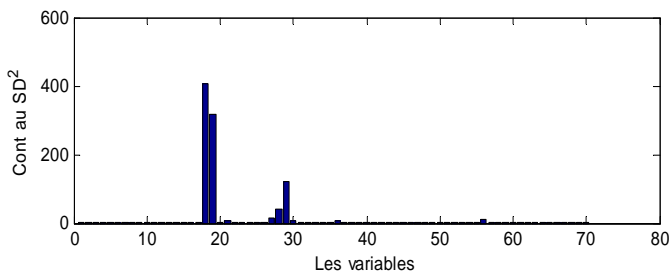


Fig. 10
LA CONTRIBUTION AU SD^2 POUR L'OBSERVATION 2755

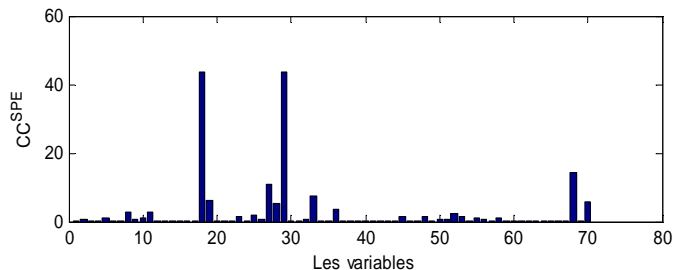


Fig. 11
LA CONTRIBUTION CLASSIQUE À LA SPE POUR L'OBSERVATION 2755

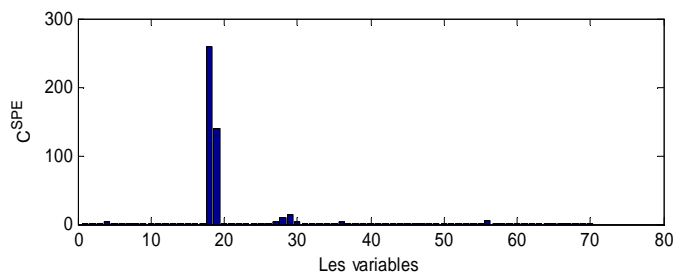


Fig. 12
LA NOUVELLE CONTRIBUTION À LA SPE POUR L'OBSERVATION 2755

à des résultats encourageants et prometteurs ce qui a satisfait l'industriel. Ces résultats probants nous ont permis d'atteindre notre objectif notamment la réduction du taux de *hold*.

Nous avons proposé ensuite une nouvelle méthode de localisation basée sur le calcul de la contribution dans l'espace résiduel de l'ACP. Les résultats obtenus sont convainquants et dans un cadre de perspective de ce travail, nous allons faire valider ces résultats par les ingénieurs de ST-Microelectronics.

RÉFÉRENCES

- [1] D. Wang and J.A. Romagnoli, « Robust multi-scale principal components analysis with applications to process monitoring, » *Journal of Process Control*, vol. 15, no. 8, pp. 869–882, 2005.
- [2] L. Lebart, A. Morineau et M. Piron, *Statistique Explorative Multidimensionnelle*, Dunod, Paris, France, 2004.
- [3] S. Buisine et J.C. Martin, « L'étude de Corpus par Analyse en Composantes Principales, » *Second Workshop sur les Agents Conversationnels Animés : WACA'2006*, Toulouse, France, 2006.
- [4] T. Kourti, P. Nomikos and J.F. MacGregor, « Analysis, monitoring and fault diagnosis of batch process using multiblock and multiway PLS, » *Journal of Process Control*, vol. 5, no. 4, pp. 277–284, 1995.
- [5] S. Verboven and M. Hubert, « LIBRA : a MATLAB Library for Robust Analysis, » *Chemometrics and Intelligent Laboratory Systems*, vol. 75, no. 2, pp. 127–136, 2005.
- [6] M. Hubert, P.J. Rousseeuw and K.V. Branden, « ROBPCA : A New Approach to Robust Principal Component Analysis, » *Technometrics*, vol. 47, no. 1, pp. 64–79, 2005.
- [7] M.F. Harkat, G. Mourot et J. Ragot, « Différentes méthodes de localisation de défauts basées sur les dernières composantes principales, » *Conférences Internationale Francophone d'Automatique*, Nantes, France, pp. 229–234, 2002.
- [8] J.A. Westerhuis, S.P. Gurden and A.K. Smilde, « Generalized contribution plots in multivariate statistical process monitoring, » *Chemometrics and Intelligent Laboratory Systems*, vol. 51, no. 1, pp. 95–114, 2000.
- [9] T. Kourti and J.F. MacGregor, « Multivariate SPC Methods for Processes and Product Monitoring, » *Journal of Quality Technology*, vol. 28, no. 4, pp. 409–428, 1996.
- [10] S. Engelen, M. Hubert and B.K. Vanden, « A Comparison of Three Procedures for Robust PCA in High Dimensions, » *Austrian Journal of Statistics*, vol. 34, no. 2, pp. 117–126, 2005.
- [11] The MathWorks. *Statistics Toolbox User's guide For Use with MATLAB*. The MathWorks Inc, Natick, USA, 2002.
- [12] H.H. Hotelling, « Multivariate Quality Control, illustrated by the Air testing of Sample Bombsights, » *Techniques of Statistical Analysis*, McGraw-Hill, New York, pp. 111–184, 1947.
- [13] H.D. Lin, « Automated visual inspection of ripple defects using wavelet characteristic based multivariate statistical approach, » *Image and Vision Computing*, vol. 25, no. 11, pp. 1785–1801, 2007.
- [14] B. Ananou, A. Boubouzoul, M. Ouladsine, H. Noura, F. Gasnier and J. Pinaton, « Aid decision and system based on statistical multivariate approach : applied to wafer manufacturing, » *AEC/APC Symposium XVIII*, Westminste, Colorado, 2006.
- [15] T. Kourti, « Application of latent variable methods to process control and multivariate statistical process control in industry, » *International Journal of Adaptive Control and Signal Processing*, vol. 19, no. 4, pp. 213–246, 2005.
- [16] J.-M. Lee, C. Yoo and I.-B. Lee, « Statistical process monitoring with independent component analysis, » *Journal of Process Control*, vol. 14, no. 5, pp. 467–485, 2004.