International Journal of Mechanical, Industrial and Aerospace Sciences
ISSN: 2517-9950
Vol:6, No:12, 2012

Yield Prediction using Support Vectors based Under-Sampling in Semiconductor Process

Sae-Rom Pak, Seung Hwan Park, Jeong Ho Cho, Daewoong An,
Cheong-Sool Park, Jun Seok Kim, Jun-Geol Baek

Abstract—It is important to predict yield in semiconductor test process in order to increase yield. In this study, yield prediction means finding out defective die, wafer or lot effectively. Semiconductor test process consists of some test steps and each test includes various test items. In other world, test data has a big and complicated characteristic. It also is disproportionably distributed as the number of data belonging to FAIL class is extremely low. For yield prediction, general data mining techniques have a limitation without any data pre-processing due to eigen properties of test data. Therefore, this study proposes an under-sampling method using support vector machine (SVM) to eliminate an imbalanced characteristic. For evaluating a performance, randomly under-sampling method is compared with the proposed method using actual semiconductor test data. As a result, sampling method using SVM is effective in generating robust model for yield prediction.

Keywords—Yield Prediction, Semiconductor Test Process, Support Vector Machine, Under Sampling

I. Introduction

AMONG various manufacturing industries, semiconductor industry along with other industries has been intensively developed and demand is growing steadily [1]. Many semiconductor manufacturing companies have managed for production cycle, process variability, yield and quality management and etc. in order to produce better products among companies [2]. Specially, yield management of semiconductor process is one of the essentially important requirements for cost reduction and competing with companies. To produce high quality product, the semiconductor manufacturing companies carry out yield management. Therefore, companies make many efforts that check Lot history, regular equipment maintenance, yield prediction and etc. for yield management. However, it is difficult to predict yield in management due to converging data from many complicated manufacturing system during long term. Even if the yield in final step is high levels which consist pass, each yield in several test steps is relatively lower. Because of about 99.9% in final yield, also, it is hard to predict yield consisted of pass or fail through general statistical methods for semiconductor data. Therefore, each step-by-step test result must be improved about yield prediction performance increasingly. Generally, the semiconductor manufacturing processes are composed of Fabrication process, Test 1, Assembly process and Final test in Fig.1.

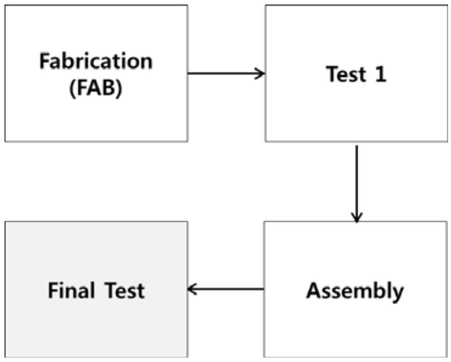

Fig. 1: Gneral flow in Semiconductor process

Many semiconductor works about yield prediction have been considered on Test1 (see Fig.1). Ciciani and Jazeolla [3] proposed new calculation method for yield calculation in chips using statistical distribution. But, the shortcoming is that only one variable is considered without the other variables. Recently, Daewoong [4] proposed efficient classification methodology that Stepwise-SVM (SSVM) is accurately prediction to yield how to adjusting parameters for each step. In addition, Verdier and Ferreira [5] proposed fault detection method using k-nearest neighbor rule not the Euclidean distance but the Mahalanobis distance. Kittisak and Nittaya [6] proposed fault detection from over-sampling method duplicating major class based decision tree algorithm. In spite of these studies, the semiconductor processing data still generate problems about yield prediction of fault data.

The data that is generated from a semiconductor processing has many variables, highly voluminous characteristic. Data category consists of binary class pass or fail. Almost data is belonged as pass. Therefore, yield prediction about fail is difficult because yield of semiconductor data regards almost results as pass. Due to these characteristics, two major problems occur when yield is predicted through machine learning or statistical methods. The first problem is one-sided imbalance data [7]. This phenomenon is called imbalanced problem. Problem of imbalanced data lead to obstacle for machine learning algorithms that decide to classify data [8]. The second problem is voluminous data. Semiconductor production unit is comprised of a Lot that is made up 25 wafers and each wafer is composed of about 1000–2000 dies. Therefore, the

Sae-Rom Pak, Seung Hwan Park, Cheong-Sool Park and Jun Seok Kim are with the School of Industrial Management Engineering, Korea University Anam-dong, Seongbuk-gu, Seoul, 136-713, Republic of Korea

Jeong Ho Cho and Daewoong An are with SKhynix INC., 2091, Gyeongchung-daero, Bubal-eub, Icheon-si, Gyeonggi-do, Republic of Korea

Jun-Geol Baek is an associate professor in the School of Industrial Management Engineering, Korea University, Seoul, 136-713, Republic of Korea (Corresponding author to provide phone: 82-2-3290-3396; fax: 82-2-929-5888; email: jungeol@korea.ac.kr).

number of total die in a Lot is about 50000 (2000×25) and the semiconductor process produce many Lots during one day. Consequently, semiconductor data is big, so it is difficult to deal with this data.

To solve the problems of imbalanced and voluminous data, this paper proposes that each wafer of a Lot creates model for under-sampling method. Under-sampling in each wafer is implemented by extracting Support Vectors of Support Vector Machine(SVM). Then, extracted Support Vectors of each wafer are set to merge new dataset. Merged dataset creates new one robust model.

II. Support Vector Machine (SVM)

Support Vector Machine (SVM) algorithm is well known two class classifier by Vapnik [9]. SVM has shown outstanding performance as effective classifier in several areas like face recognition [10], text categorization [11], [12] and etc.

Purpose of SVM is to find optimal hyper plane which can be well separated. Furthermore, this algorithm solves a non-linear problem based on kernel through high dimension feature space as mapping ϕ of the input data [9]. Then, a function $K(\cdot)$ return to the inner product $\langle\phi(x), \phi(x')\rangle$ with raw two points x, x'. In other words, this approach is called Kernel Trick [13].

Fig.2 is shown mapping from non-linear data to kernel space. Kernel function is defined as follow (1).

$$K(\cdot) = \langle\phi(x), \phi(x')\rangle \tag{1}$$

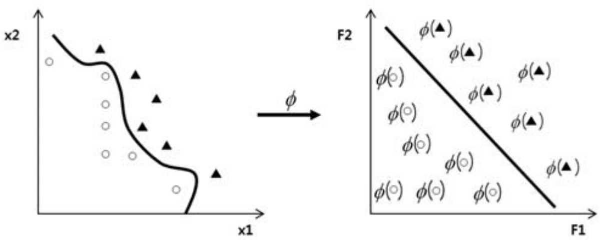

Fig. 2: Mapping to kernel from non-linear data

Kinnds of Kernel function are consisted as Gaussian Radia, polynormial, sigmoid and etc. Generally, RBF function (2) and polynomial function (3) have been frequently used and using RBF kernel function in this paper.

- *The Gaussian Radial Basis Function(RBF) kernel function*

$$K(\cdot) = \exp(\|x - x'\|^2/2\sigma^2) \tag{2}$$

- *The polynomial kernel function*

$$K(\cdot) = (\langle x, x'\rangle + 1)^p, \qquad p > 0 \tag{3}$$

In addition, non-linear SVM includes the constraints on the slack variable ξ to consider causing misclassification for solving non-linear problem. Equation (4) is represent constrain which minimize the reciprocal of margin.

$$y_i(\mathbf{w}'x_i + b) \geq 1 - \xi_i, \forall i = (1, 2, \ldots, N) \tag{4}$$

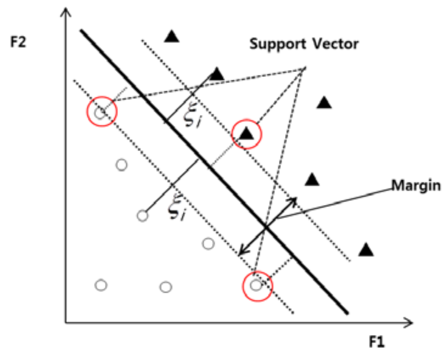

Fig. 3: Support Vector Machine in non-linear

If input data composed of m dimension and non-linear separable N data, It is necessary for N variables consisted of $\mathbf{w_1}, \mathbf{w_2}, \ldots, \mathbf{w_m}, b, \xi_1, \xi_2, \ldots_N$. To find hyperplane for minimum error rate, a function $\phi(\xi)$ is defined as (5).

$$\phi(\xi) = \sum_{i=1}^{N} l(\xi_i - 1), \qquad \text{if } l(\xi_i) = \begin{cases} 0, & \xi_i \leq 0 \\ 1, & \xi_i > 0 \end{cases} \tag{5}$$

Because it is difficult to solve the problem in this way, this problem has to estimating equation as upper bound. Estimated equation is as following (6).

$$\phi'(\xi) = \sum_{i=1}^{N} \xi_i \tag{6}$$

Therefore, the objective function is (7) including slack variable.

$$\begin{aligned} &\min L(W, \xi) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N} \xi_i \\ &st. \begin{cases} y_i(\mathbf{w}'x_i + b) \geq 1 - \xi_i, & \forall i \\ \xi_i \geq 0, & \forall i \end{cases} \end{aligned} \tag{7}$$

Where C is the cost parameter, it represents trade-off between misclassification rate and performance. Usually, misclassification rate is decreasing if value of C is increasing. On the contrary, the more C is increasing, the more minimum distance is maximizing. Complication of solving problem also decreases. The SVM objective function can be transformed into Lagrangian optimization method in (8) including 'Lagrangian multiplier α_i'.

$$\begin{aligned} L(\mathbf{w}, b, \alpha) = \; & \tfrac{1}{2}\mathbf{w}'\mathbf{w} + C\textstyle\sum_{i=1}^{N} \xi_i \\ & - \textstyle\sum_{i=1}^{N} \alpha_i[y_i(\mathbf{w}'x_i + b) - 1 + \xi_i] \\ & - \textstyle\sum_{i=1}^{N} \mu_i \xi_i, \\ & \alpha_i \geq 0, \mu_i \geq 0 \end{aligned} \tag{8}$$

To solve (8), using Quadratic Programming(QP)(9) computes Lagrangian multiplier α_i as satisfying Karush-Kuhn-Tucker(KKT) conditions. KKT conditions are necessary and sufficient condition to find solution in Convex Optimization problem.

- KKT(Karush-Kuhn-Tucker) conditions

1) $\frac{\partial L(\mathbf{w},b,\alpha_i)}{\partial \mathbf{w}} = \mathbf{w} - \sum \alpha_i y_i x_i = 0, \quad \mathbf{w} = \sum \alpha_i y_i x_i$

2) $\frac{\partial L(\mathbf{w},b,\alpha_i)}{\partial \mathbf{w}} = -\sum \alpha_i y_i = 0$

3) $\alpha_i \geq 0, \qquad i = 1, 2, \ldots, N$

4) $\alpha_i[y_i(\mathbf{w}'x_i + b) - 1 + \xi_i] = 0, \qquad i = 1, 2, \ldots, N$

- Objective function using QP

$$\max \quad Q(\alpha) = \sum \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$
$$st. \sum_{i=1}^{N} \alpha_i y_i = 0,$$
$$0 \leq \alpha_i \leq C, \qquad i = 1, 2, \ldots, N \tag{9}$$

According to last condition of KKT conditions, Equiation (10) is induced. Decision boundary and decision function is expressed as (11),(12).

$$b = \frac{i - \xi_i}{y_i} - \mathbf{w}'x_i, \qquad b = 1 - \xi_i - \mathbf{w}'x, (y_i = 1) \tag{10}$$

$$\mathbf{w}'x + b = \sum_{i=1}^{n'} \alpha_i y_i K(x, x') + b \tag{11}$$

$$f(x) = \sum \alpha_i y_i K(x, x') + b \tag{12}$$

In addition, the support vector is defined according to the range of α_i, which is $0 < \alpha_i < C$.

III. Support Vectors based Under-sampling

Typically, most algorithms affect learning performance due to imbalanced characteristic. So, it must be generally dealt with methods like over-sampling or under-sampling after preprocessing. Over-sampling is replicative in the minor class while this case occurs up problem such as overlapping data of minor class. Under-sampling eliminates data belonging to majority class as adjusting the number of minor class data set [7]. Therefore, to solving problem in voluminous and imbalanced data at the same time, the proposed method is under-sampling technique using Support Vectors of majority class. The SVM has been proved to effectively classify data from specially imbalanced data set [10], [11], [12]. However, when kernel matrix is calculated in the kernel space, the problem occurs to calculation which is $m \times m$ matrix (m is the number of observation) in the real. As a result, it is handled to big or voluminous data such as semiconductor test data. Proposed method is following steps.

Step1. Each divided and imbalanced data is learned through SVM algorithm.

Step2. It generates support vectors of major class corresponding to minor support vector data in SVM model of each data.

Step3. The sampled data create new dataset which is composed of support vector from each divided data. And new data is learned through SVM again.

Step4. SVM model using new data consisted of support vectors is constructed to predict yield.

Before SVM is learned for extracting Support Vectors in Step1,the number of both classes data is balanced by adjusting parameters of SVM. Extrainting data as Support vectors has characteristic that almost minority data is fixed as Support Vectors. It can define meaningful data on majority class of Support Vectors corresponding minority class. In Step 4, SVM parmeter is provided through heuristic method for creating new model likwise Step 2.

The reason these process is conducted predicts yield of semiconductor model through robust model.

IV. Experiment

In this experiment, actual data from semiconductor test process are used and consists of continuous input variables and binary output variable. Data preprocessing such as outlier detection, normalization and variable is conducted to increase performance of model. Information with regard to data pre-processing are provided from field engineers.

As applying proposed method, Support Vectors are extracted from corresponding minority class from each wafer. SVM parameter for extracting Support Vectors is fixed in all same sigma(0.2×10^{-15}) in wafers. This value is maintained as uniform Support Vectors of one to one ratio in class (pass and fail). The reason sigma value is very small is that SVM using RBF kernel is determined from sigma and Support Vectors are influenced by sigma.

Extracted Support Vector set are regarded as new data set representing each wafer. New data set is learned for creating new model through 'nu-SVM'. The parameter value of 'nu-SVM' decides optimal value through heuristic method. On the basis of the above processing, one wafer is test set and the others is train set. Experiment is conducted with 25th cross validation.

A. Test Measure

Generally, evaluation of performance is accuracy . But imbalanced data consisted of 99% major class and 1% minor class accuracy is close to 99% accuracy. Therefore, not total accuracy but each accuracy (see TABLE I). Also, positive and Negative class define Pass and Fail.

– True negative rate (TNR) : $\frac{TN}{FP+TN}$
It is true negative case which correctly classified.

– True positive rate (TPR) : $\frac{TP}{TP+FN}$
It is true positive case which correctly classified.

– False negative rate (FNR) : $\frac{FN}{TP+FN}$
It is false negative case which incorrectly classified.

– False positive rate (FPR) : $\frac{FP}{FP+TN}$
It is false positive case which incorrectly is classified.

TABLE I
Confusion Matrix for Evaluating Performance

	Positive Prediction(Pass)	Negative Prediction(Fail)
Actual Positive class (Pass)	True Positive(TP)	False Negative(FN)
Actual Negative class(Fail)	False Positive(FP)	True Negative(TN)

Another measure is Geometric Mean (GM) for considering each accuracy of 'True Positive' and 'True Negative'. GM is defined as $\sqrt{TPR \cdot TNR}$ [14]. It is to maximize for balanced correct accuracy between positive and negative class.

GM measure means closer according to 1, it represents good performance. On the contrary, value according to 0 is bad performance. This paper uses TPR, TNR and GM.

B. Results

TABLE II
Performance Comparison with two methods

	Support Vectors based Under-sampling			Randomly Under-sampling		
nu=0.007	SV_TNR	SV_TPR	SV_GM	R_TNR	R_TPR	R_GM
result1	0.741	0.538	0.632	0.143	0.760	0.330
result2	0.866	0.367	0.564	0.176	0.847	0.387
result3	0.200	0.322	0.254	0.143	0.795	0.337
result4	0.448	0.379	0.412	0.600	0.811	0.697
result5	0.330	0.253	0.289	0.333	0.966	0.567
result6	0.258	0.475	0.350	0.429	0.929	0.631
result7	0.279	0.548	0.391	0.500	0.861	0.656
result8	0.705	0.574	0.636	0.000	0.824	0.000
result9	0.797	0.627	0.707	0.500	0.868	0.659
result10	0.554	0.794	0.663	0.143	0.888	0.356
result11	0.239	0.457	0.330	0.000	0.933	0.000
result12	0.904	0.536	0.696	0.091	0.865	0.280
result13	0.742	0.730	0.736	0.167	0.833	0.373
result14	0.159	0.460	0.270	0.400	0.721	0.537
result15	0.797	0.450	0.599	0.000	0.948	0.000
result16	0.182	0.761	0.372	0.182	0.964	0.419
result17	0.149	0.568	0.290	0.286	0.851	0.493
result18	0.747	0.461	0.587	0.091	0.867	0.281
result19	0.165	0.588	0.311	0.143	0.862	0.351
result20	0.212	0.541	0.339	0.154	0.834	0.358
result21	0.722	0.495	0.598	0.083	0.906	0.275
result22	0.115	0.713	0.286	0.267	0.847	0.475
result23	0.890	0.364	0.570	0.083	0.878	0.271
result24	0.787	0.506	0.631	0.222	0.893	0.446
result25	0.492	0.454	0.472	0.125	0.943	0.343
average	0.499	0.518	0.479	0.210	0.868	0.381

Support Vectors based under-sampling has usually better performance than randomly under-sampling. Accuracy value of TNR and TPR has generally the opposite tendency each other. For example, if TPR value is high, TNR value is low. However, proposed method shows similar performance both SV_TNR and SV_TPR than results of randomly under-sampling method (show TABLE II). However, R_TNR and R_TPR are greatly one side for R_TPR in randomly under-sampling method. In other words, random sampling misclassifies minority class (fail).

In TABLE II, each result value of test9 and test13 represents specially the best performance which accounts for about 80%, 63% and 74%, 73% in SV_TNR and SV_TPR. Each SV_GM of test9 and test13 accounts for 71%, 74% and average SV_GM is about 10% higher than R_GM. As a result, Support Vectors based Under-sampling method is better than Randomly Under-sampling method. It has also shown uniformly maintaining performance except for some cases.

Fig.4 is represented for each boxplot between GM of under-sampling using support Vector and randomly under-sampling. Boxplot shows difference of SV_GM and R_GM

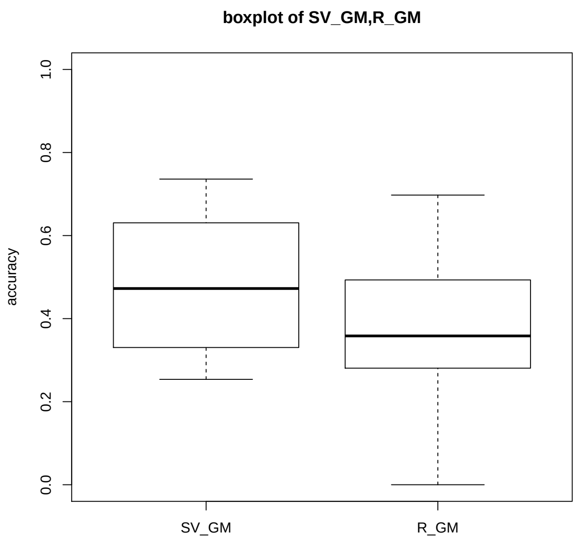

Fig. 4: Boxplot of Support Vector sampling and randomly under-sampling(GM)

which consider both TNR and TPR.

Fig.5 shows boxplot of TNR and TPR from Support Vectors based Under-sampling and Randomly Under-sampling. SV_TNR and SV_TPR of Fig.5 boxplot maintain generally similar value. That is to say that Randomly Under-sampling method is not well to predict minority class.

V. Conclusion

Yield management of the semiconductor manufacturing process is important. As generated data have a big and complicated characteristic, it has difficult predicting yield. This study proposed sampling method to improve a performance of yield prediction. Proposed method is under-sampling based

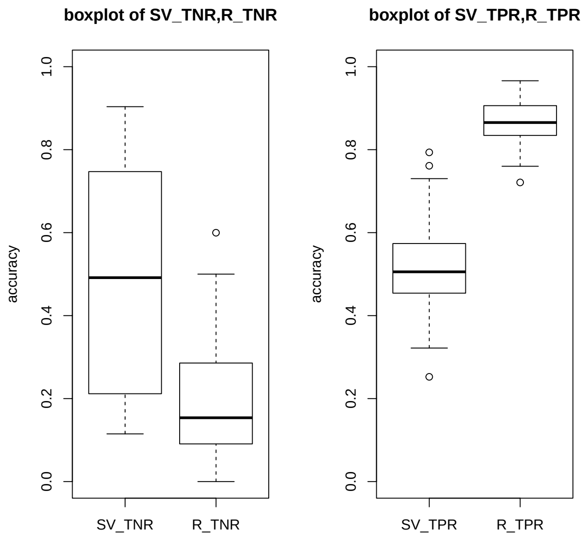

Fig. 5: Boxplot of TNR and TPR

on Support Vectors. The Support Vectors from the SVM are regarded as sample of each wafer. The new model that uses by merging Support Vectors has a superior performance than random sampling method.

This study can conclude meaningful sampling method through experiment results. Under sampling of the majority class as using Support Vectors reflects data characteristic. It also makes the conclusion to be robust model as similar test results of TNR and TPR through trials of 25 tests. This method need to apply further additional data in order to increase accuracy of proposed model.

Acknowledgment

This research was supported by the MKE (Ministry of Knowledge Economy), Korea, under the IT R&D Infrastructure Program supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2012-(B1100-1101-0002)).

References

[1] K. Kim, C.-G. Hwang, and J. G. Lee, "Dram technology perspective for gigabit era," *Electron Devices, IEEE Transactions on*, vol. 45, no. 3, pp. 598 –608, mar 1998.
[2] P. Burggraaf, "New roadmap unveiled-2000 begins with a revised industry roadmap-a revised 1999 international technology roadmapfor semiconductors identifies future challenges to the evolution of semiconductor," *Solid State Technology*, vol. 43, no. 1, pp. 31–48, 2000.
[3] B. Ciciani and G. Iazeolla, "A markov chain-based yield formula for vlsi fault-tolerant chips," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 10, no. 2, pp. 252–259, 1991.
[4] D. An, H. Ko, T. Gulambar, J. Kim, J. Baek, and S. Kim, "A semiconductor yields prediction using stepwise support vector machine," in *Assembly and Manufacturing, 2009. ISAM 2009. IEEE International Symposium on*. IEEE, 2009, pp. 130–136.
[5] G. Verdier and A. Ferreira, "Adaptive mahalanobis distance and¡ formula formulatype=," *Semiconductor Manufacturing, IEEE Transactions on*, vol. 24, no. 1, pp. 59–68, 2011.
[6] K. Kerdprasop and N. Kerdprasop, "A data mining approach to automate fault detection model development in the semiconductor manufacturing process."
[7] N. Chawla, K. Bowyer, L. Hall, and W. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *arXiv preprint arXiv:1106.1813*, 2011.
[8] X. Liu, J. Wu, and Z. Zhou, "Exploratory undersampling for class-imbalance learning," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 39, no. 2, pp. 539–550, 2009.
[9] V. Vapnik, *The nature of statistical learning theory*. springer, 1999.
[10] E. Osuna, R. Freund, and F. Girosit, "Training support vector machines: an application to face detection," in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*. IEEE, 1997, pp. 130–136.
[11] S. Dumais, J. Platt, D. Heckerman, and M. Sahami, "Inductive learning algorithms and representations for text categorization," in *Proceedings of the seventh international conference on Information and knowledge management*. ACM, 1998, pp. 148–155.
[12] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," *Machine learning: ECML-98*, pp. 137–142, 1998.
[13] B. Schlkopf and A. Smola, "Learning with kernels: Support vector machines, regularization, optimization, and beyond," 2002.
[14] R. Barandela, J. Sánchez, V. García, and E. Rangel, "Strategies for learning in class imbalance problems," *Pattern Recognition*, vol. 36, no. 3, pp. 849–851, 2003.