Adaptive Weighted Expected Improvement With Rewards Approach in Kriging Assisted Electromagnetic Design

Song Xiao, Mihai Rotaru, and Jan K. Sykulski, *Fellow, IEEE*

Electronics and Computer Science (ECS), University of Southampton, Southampton, SO17 1BJ, U.K.

The paper explores kriging surrogate modelling combined with expected improvement approach for the design of electromagnetic devices. A novel algorithm based on the concept of rewards is proposed, tested and demonstrated in the context of TEAM Workshop Problem 22. Balancing exploration and exploitation is emphasized and robustness of the design considered.

Index Terms—Computational electromagnetics, design optimization, pareto optimization, topology.

I. INTRODUCTION

THE paper builds on our previous publication [1] where Adaptive Weighted Expected Improvement (AWEI) was first introduced and the concept of rewards [2] implemented and tested using the Schwefel test function [3]. In this paper the technique is investigated further, extended to real design problems and illustrated using the TEAM Workshop Problem 22 [4]. The algorithm exploits kriging [1], [5] as an efficient tool to balance exploration with exploitation.

II. BALANCING EXPLORATION AND EXPLOITATION

A. Kriging

As a kind of regression model, kriging [5] is able to exploit the spatial correlation of data in order to predict the shape of the objective function based only on limited information. Moreover, it can estimate the accuracy of this prediction, which may be extremely helpful in assisting the main decision of any optimization process, namely where the next evaluation point (vector) should be located.

Kriging exploits the spatial correlation of data in order to build interpolation; therefore the correlation function is a critical element. This relies on the linear regression model (1) and the Gaussian correlation model (2)

$$\hat{y}(x) = \sum\nolimits_{k=1}^{m} \beta_k f_k(x) + \varepsilon(x) \tag{1}$$

$$R(\varepsilon(x^i), \varepsilon(x^i)) = \prod\nolimits_{k=1}^{n} e^{-\theta_k |x_k^i - x_k^j|^{p_k}}. \tag{2}$$

where the global function $\sum_{k=1}^{m} \beta_k f_k(x)$ and an additive Gaussian noise $\varepsilon(x)$ are integrated to the predicted value $\hat{y}(x)$ of the objective function; θ_k is the correlation amongst the data in k-direction and p_k determines the 'smoothness' of (2). The most popular correlation function is given by the Gauss model where the value of p_k is simply taken as equal to 2. For a given set of data, the maximum likelihood estimation optimizes the value of θ and then the correlation model is brought into the regression model to evaluate the function with the best linear unbiased predictor [5].

B. Utility Functions

The 'expected improvement' utility function [6], based on a potential error—the 'Mean Square Error' (MSE) produced by the kriging model—is commonly used to select multiple designs for further evaluation. Here the optimization problem is viewed as minimizing the objective function. The Expected Improvement (EI) function is defined as

— if $s(x) > 0$

$$\text{EI} = (f_{\min} - \hat{y}(x))\Phi\left(\frac{f_{\min} - \hat{y}(x)}{s(x)}\right) + s(x)\emptyset\left(\frac{f_{\min} - \hat{y}(x)}{s(x)}\right) \tag{3}$$

— if $s(x) = 0$

$$\text{EI} = 0. \tag{4}$$

where $\hat{y}(x)$ is the value of objective function predicted by the kriging model, and $s(x)$ is the root mean squared error in this prediction. The first term of (3), including Gaussian density, favors "exploitation"—searching the most promising regions (high confidence); while the second term, containing the Gaussian distribution, prefers "exploration"—searching the regions that have high uncertainty. It has been found [2] that in practical cases exploration performs dramatically better in terms of finding the global optimum, whereas exploitation often causes the kriging model to stop around a local minimum. The termination criteria used by the kriging model are often based on finding repeatedly the same sampling point within prescribed tolerance, thus balancing exploration and exploitation is vital. The EI algorithm distributes the weights equally between the two terms and can be seen as a fixed compromise between exploration and exploitation. In order to study the effect of different weights, the Weighted Expected Improvement (WEI) [6] was derived from EI by adding a tuneable weighting parameter. Through a set of experiments it was shown that by changing the value of the tunable parameter the efficiency of finding the global minimum can be affected. However, the optimal value of the weight is hard to find and usually problem dependent. Hence a modified strategy is required to design a more intelligent method which could guide itself automatically through the whole searching procedure using an iterative process.

C. Adaptive Weighted Expected Improvement With Reward Approach (AWEI)

Two algorithms inspired by reinforcement learning [7], the Adaptive Weighted Expected Improvement (AWEI) and the Surrogate Model based Weighted Expected Improvement approach utilizing the concept of rewards [2] (SMWEI), were proposed previously [1] to make the process of tuning weights
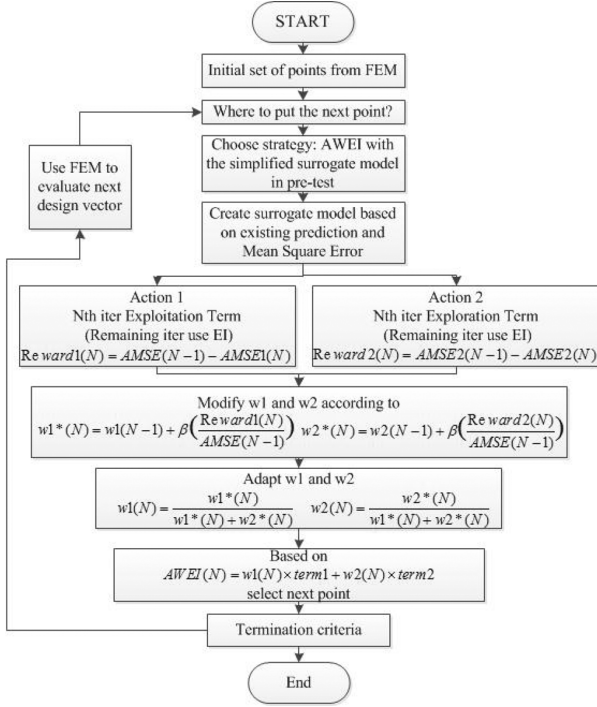
Fig. 1.   The decision-making chart for balancing exploration and exploitation using AWEI algorithm.
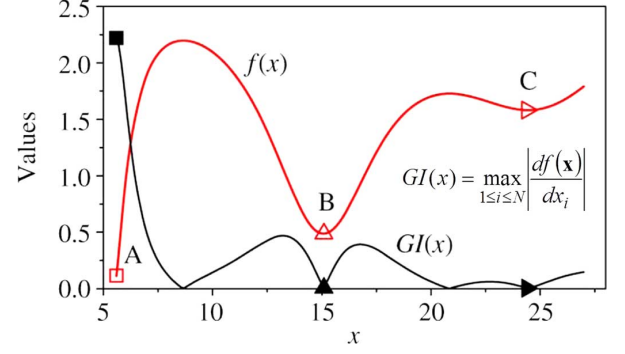

Fig. 2.   The single-variable analytic function [8], [9]. The gradient index is defined as the maximum component of the gradient of the objective function with respect to design variables [8].
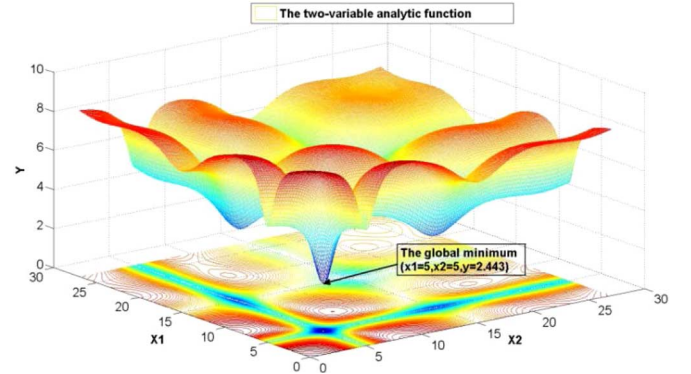

Fig. 3.   A two-variable objective function.

more intelligent and self-guiding. The decision-making chart of AWEI is shown in Fig. 1. Compared with the original [1] this version is enhanced by the addition of a pre-test stage during which a "cheap" simplified surrogate model is applied built on a specific prediction and the potential error produced by kriging. A pair of fixed weights (one emphasizing exploration and the other exploitation) are initially set at equal values. The Mean Square Error from the kriging model is used to calculate the rewards. After comparing the rewards obtained from the pre-test, the weights are redistributed on the terms favoring exploration or exploitation. The weight with a better reward of the two is then used to feed back—via the FEM module—into the main iterative loop of the design process.

Although finding the global optimum is often sufficient, this may not be enough when a robust design is considered, as information about the shape of the objective function and position and shape of local minima is also relevant. The quality of the approximation of the objective function is the most important prerequisite when the robustness of the solution is based on a kriging model rather than the real function. In the next section the performance of the AWEI method will be assessed in the context of finding the global optimum concurrently with obtaining a good quality function prediction for the search space of interest.

III. NUMERICAL EXPERIMENTS

To explain the concepts we use a simple single-variable function shown graphically in Fig. 2. There are two local minima (B and C) and one global minimum (A). When considering practical factors such as uncertainty of the variables due to manufacturing tolerances, non-uniform material properties or imperfect control of operating conditions, the theoretically best-performing point A may not be a practical optimal solution. The final judgment will be influenced by the margin of the uncertainty and by how much the objective function

may change when the design variables assume limiting values. For a minor uncertainty, even under extreme (but still small) departure of the parameter from its nominal value, the objective function may still be acceptable—even if worse than at the theoretical "best"—and thus even a "sharp" optimum like A may be judged as practical. But under increased uncertainty the confidence in the final performance may be assured only by a much more "shallow" minimum, such as B, or even C. Thus the preferred practical design may be selected away from the theoretical global optimum.

Various methods of assessing the robustness have been proposed in literature. In [9], the initial optimization problem is transformed into a multi-objective optimization where both the objective function and the gradient index of this function are minimized simultaneously. We take a similar approach; however, we calculate the gradient index using the predicted value of the objective function rather than the function itself. The motivation behind such an approach is to reduce the computational effort by avoiding the costly FEM (or similar) numerical solution. For this methodology to work it is very important to have good prediction for the objective function. The quality of the prediction will now be investigated, initially with the aid of a popular test function [8], [9] defined as

$$f(x) = 10 - \sum_{i=1}^{n} \left[\frac{3.5}{1+(x_i-5)^2} + \frac{2.2}{1+(x_i-15)^2/10} + \frac{1.2}{1+(x_i-25)^2/30} \right] \quad (5)$$

in the range $5.6 \leq x_i \leq 27$. The two-variables $(n=2)$ version is plotted in Fig. 3 and will be used to assess AWEI and compare its performance against standard EI method.
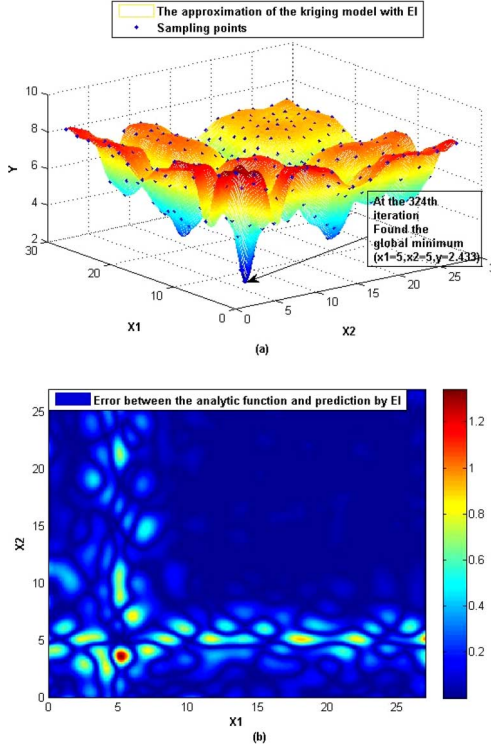
Fig. 4. Kriging assisted EI; (a) function after 324 iterations, (b) error.

Fig. 5. Kriging assisted AWEI; (a) function after 211 iterations, (b) error.

For both the EI and AWEI tests, seven initial sampling points were used (as in Table I). Kriging assisted EI needs 324 iterations to find the global minimum and the predicted shape is demonstrated in Fig. 4(a), with the error calculated against the actual objective function shown in Fig. 4(b). The AWEI strategy is more efficient in finding the global minimum, requiring 211 iterations (Fig. 5), but the final approximation is not as good as that resulting from EI as the number of available points is less. Nonetheless, in some regions, e.g., $5 \leq x_1 \leq 27, 2.5 \leq x_2 \leq 7.5$ or $2.5 \leq x_1 \leq 7.5, 5 \leq x_2 \leq 27$, AWEI outperforms EI (Fig. 5(b)).

IV. RESULTS FOR TEAM WORKSHOP PROBLEM 22

The full description of the TEAM 22 Benchmark problem (superconducting magnetic energy storage system) may be found in [10]. The goal of the optimization task is to achieve the stored energy of $E_{\text{ref}} = 180$ MJ with a minimal stray field B_{stray}. The objective function is defined as

$$\text{OF} = \frac{B_{\text{stray}}^2}{B_{\text{norm}}^2} + \frac{|E - E_{\text{ref}}|}{E_{\text{ref}}} \qquad (6)$$

where $B_{\text{norm}} = 3 \, \mu\text{T}$ and $B_{\text{stray}}^2 = (\sum_{i=1}^{22} |B_{\text{stray},i}|^2)/(22)$, subject to some geometrical and "quench" constraints. We consider here the 3 parameter problem, although initially—for the purpose of demonstrating typical shapes of the objective function—one of the variables is fixed ($R_2 = 3.08$ m), while d_2 and h_2 varied. One of the challenges in creating the kriging correlation matrix is the "combinatorial explosion" when many parameters are used [11] and we propose to use a successive "zoom in" strategy to cope with this issue; a two-stage version implemented in the context of the TEAM Problem 22 is described in Table II.

The initial sampling points would normally be selected using the Latin Hypercube [12]; in our tests we fixed the positions at
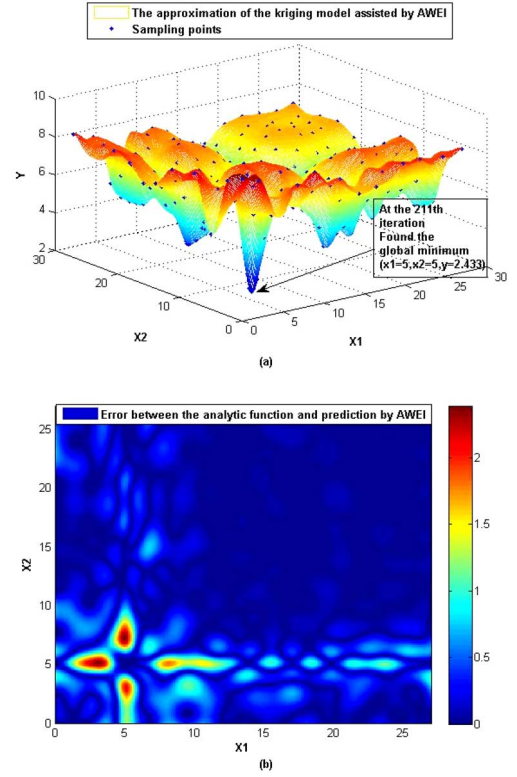
TABLE I
INITIAL SAMPLING POINTS

x_1	x_2	y	x_1	x_2	y
10	7	7.9966	20	7	7.6024
10	10	8.1913	7	13	7.0688
5.7	10	6.4292	25	20	7.2927
13	7	7.0688			

TABLE II
THE 'ZOOM IN' STRATEGY USING KRIGING ASSISTED EI FOR TEAM 22 TESTS

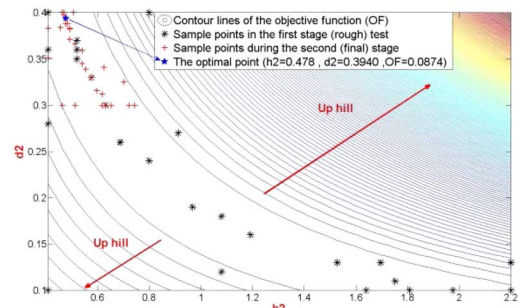|		The first stage		The second stage	
		$h_2(m)$	$d_2(m)$	$h_2(m)$	$d_2(m)$
Test range	Min	0.408	0.1	0.464	0.35
	Max	2.2	0.4	0.576	0.4
Step size		0.056	0.01	0.007	0.001

Fig. 6. The two-parameter SMES problem tested by kriging with EI.

($h_2 = 0.744$ m, $d_2 = 0.13$ m), ($h_2 = 1.304, d_2 = 0.22$ m) ($h_2 = 1.64$ m, $d_2 = 0.40$ m) and ($h_2 = 2.088$ m, $d_2 = 0.37$ m) to facilitate comparisons. Kriging with EI required 25 iterations to complete the first "rough" stage, and further 29 iterations to
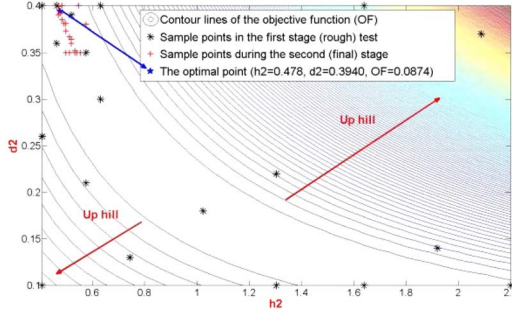
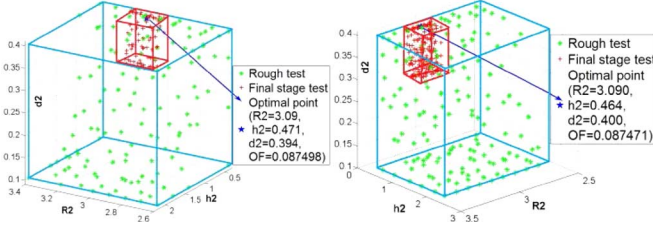Fig. 7. The two-parameter SMES problem tested by kriging with AWEI.

Fig. 8. Kriging applied to three-parameter SMES problem (a) EI, (b) AWEI.

TABLE III
THE THREE-PARAMETER KRIGING ASSISTED AWEI FOR TEAM 22 TESTS

		The first rough stage			The second stage		
		$R_2[m]$	$h_2[m]$	$d_2[m]$	$R_2[m]$	$h_2[m]$	$d_2[m]$
Test range	Min	2.6	0.408	0.1	3.0	0.408	0.3
	Max	3.4	2.2	0.4	3.2	0.744	0.4
Step size		0.1	0.056	0.01	0.01	0.007	0.001

find the best-performing solution ($h_2 = 0.478$ m, $d_2 = 0.394$ m, OF = 0.0874 (Fig. 6). For kriging assisted AWEI the relevant iteration numbers are 21 and 23 (Fig. 7). The 'history' of the EI and AWEI strategies may be followed on the two figures.

The initial settings for the 3 parameter test are presented in Table III. For kriging with EI, 79 sampling points were created in the first stage, with further 132 points in the zoomed-in region; in the AWEI case, 156 sampling points were followed by 167 points in the second stage (Fig. 8). The total numbers quoted include the initial sampling points needed by the kriging model. It is no longer possible to conveniently display the objective function itself for three parameters.

To demonstrate the advantages of the proposed methods over other well-known stochastic algorithms, Table IV has been compiled using available published data [19]. It is clear that kriging assisted EI and AWEI both offer significant advantages in terms much reduced number of computationally expensive function calls to achieve required level of accuracy. However, those gains are somewhat offset by the need to create the kriging correlation matrix. Finally, the robustness of the design requires further consideration and will be studied in the context of more relevant test cases. A particular challenge is to select test functions representative of the problems encountered in electromagnetic design, although the prime concern remains to keep the number of function calls low.

V. CONCLUSION

Exploration and exploitation in the design algorithm are controlled via a kriging model, the Weighted Expected Improvement strategy and the use of "rewards," with robust design also

TABLE IV
PERFORMANCE COMPARISON OF DIFFERENT ALGORITHMS

Algorithm	$R_2(m)$	$d_2(m)$	$h_2/2(m)$	OF	No. of Iterations
GA	3.040	0.386	0.240	0.134	2400
HuTS	3.080	0.380	0.246	0.089	3821
ITS	3.100	0.388	0.240	0.098	1824
SA	3.078	0.390	0.237	0.098	5025
NTS	3.080	0.370	0.254	0.089	1800
PBIL	3.110	0.421	0.241	0.101	3278
EI(Kriging)	3.090	0.471	0.197	0.0875	211
AWEI(Kriging)	3.090	0.464	0.200	0.0875	323

Genetic Algorithm (GA) [13]; Tabu Search (HuTS) [14]; Improved Tabu Search (ITS) [15]; Simulated Annealing Algorithm (SA) [16]; New Tabu Search (NTS) [17]; Population-based Incremental Learning (PBIL) [18].

considered. To avoid problems associated with combinatorial explosion a "zoom in" strategy is proposed.

REFERENCES

[1] S. Xiao, M. Rotaru, and J. K. Sykulski, "Exploration versus exploitation using kriging surrogate modelling in electromagnetic design," *COMPEL*, vol. 31, no. 5, pp. 1541–1551, 2012.
[2] A. Sykulski, N. Adams, and N. Jennings, "On-line adaptation of exploration in the one-armed bandit with covariate problem," in *Proc. 9th Int. Conf. Machine Learning and Applications*, 2010.
[3] H. P. Schwefel, *Numerical Optimization of Computer Models*. New York: Wiley, 1981.
[4] TEAM Workshop Problem 22 [Online]. Available: www.compumag.org/jsite/team.html
[5] L. Lebensztajn, C. A. R. Marreto, M. C. Costa, and J.-L. Coulomb, "Kriging: A useful tool for electromagnetic device optimization," *IEEE Trans. Magn.*, vol. 40, no. 2, pp. 1196–1199, Mar. 2004.
[6] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *J. Global Optim.*, vol. 13, pp. 455–492, 1998.
[7] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press, 1998.
[8] Z. Y. Ren, M. T. Pham, M. H. Song, D. H. Kim, and C. S. Koh, "A robust global optimization algorithm of electromagnetic devices utilizing gradient index and multi-objective optimization method," *IEEE Trans. Magn.*, vol. 47, no. 5, pp. 1254–1257, May 2011.
[9] Z. Y. Ren, M. T. Pham, and C. S. Koh, "Robust global optimization of electromagnetic devices with uncertain design parameters: Comparison of the worst-case optimization methods and multi-objective optimization approach using gradient index," *IEEE Trans. Magn.*, vol. 49, no. 2, pt. 2, pp. 851–859, Feb. 2013.
[10] P. G. Alotto, U. Baumgartner, F. Freschi, M. Jaindl, A. Kostinger, Ch. Magele, and W. Repetto, in *SMES Optimization Benchmark: TEAM Workshop Problem 22* [Online]. Available: http://www.igte.tugraz.at/team22
[11] G. I. Hawe and J. K. Sykulski, "A scalarizing one-stage algorithm for efficient multi-objective optimization," *IEEE Trans. Magn.*, vol. 44, no. 6, pp. 1094–1097, Jun. 2008.
[12] T. J. Santner, B. J. Williams, and W. I. Notz, *The Design and Analysis of Computer Experiments*. Berlin: Springer, 2003.
[13] J. H. Holland, *Adaption in Natural and Artificial System*. Cambridge: MIT Press, 1975.
[14] N. Hu, "Tabu search method with random moves for globally optimal design," *Int. J. Numer. Meth. Eng.*, vol. 35, pp. 1055–1071, 1992.
[15] S. L. Ho, S. Y. Yang, G. Z. Ni, and H. C. Wong, "An improved Tabu search for the global optimizations of electromagnetic devices," *IEEE Trans. Magn.*, vol. 37, no. 5, pp. 3570–3574, Sep. 2001.
[16] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, 1983.
[17] O. Hajji, S. Brisset, and P. Brochet, "A new Tabu search method for optimization with continuous parameters," *IEEE Trans. Magn.*, vol. 40, no. 2, pp. 1184–1187, Mar. 2004.
[18] S. Y. Yang, Y. N. Bai, G. H. Zhang, and L. W. Vecchi, "An improved population-based incremental learning method for inverse problems," in *Proc. Autom. Congr.*, 2008, pp. 1–4.
[19] S. G. An, S. Y. Yang, S. L. Ho, T. Li, and W. N. Fu, "A modified Tabu search method applied to inverse problems," *IEEE Trans. Magn.*, vol. 47, no. 5, pp. 1234–1237, May 2011.