Journal of
Computational and Theoretical Nanoscience
Vol. 5, 1251–1254, 2008

The Stable Finite Element Method for Minimization Problems

B. Liu[1, 2, *] and Y. Huang[3]

[1] *FML, Department of Engineering Mechanics, Tsinghua University, Beijing 100084, China*
[2] *State Key Laboratory of Structural Analysis for Industrial Equipment, Dalian University of Technology, Dalian 116023, China*
[3] *Department of Civil and Environmental Engineering and Department of Mechanical Engineering Northwestern University, Evanston, IL 60208, USA*

The conventional finite element method is difficult to converge for a non-positive definite stiffness matrix, which usually occurs when the material displays softening behavior or when the system is near the state of bifurcation. We have developed two stable algorithms for a non-positive definite stiffness matrix, one for the direct linear equation solver and the other for the iterative solver in the finite element method for minimization problems. For a direct solver with non-positive definite stiffness matrix, energy minimization of a system with multiple degrees of freedom (DOF) is decomposed to the minimization of many 1-DOF systems, and for the latter an efficient and robust minimization method is developed to ensure that the system energy decreases in every incremental step, regardless of the positive definiteness of the stiffness matrix. For an iterative solver, the stiffness matrix is modified to ensure the convergence, and the modified stiffness matrix indeed leads to the correct solution. An example of a single wall carbon nanotube under compression is studied via the proposed algorithms.

Keywords: Atomic-Scale Finite Element Method, Minimization, Iterative Solver, Direct Solver.

RESEARCH ARTICLE

1. INTRODUCTION

Minimization (or optimization) is an important branch in scientific computation. Many minimization methods for multi-dimensional systems, such as the steepest descend method and conjugate gradient method (e.g., Ref. [1]), have been developed and widely used in molecular mechanics. Recently, Liu et al.[2, 3] proposed an atomic-scale finite element method (AFEM) to study the discrete system of atoms using the finite element method. They observed that AFEM is a more efficient minimization method than the steepest descend method and conjugate gradient method, which is briefly discussed in the following.

For a system with N nodes (or atoms in AFEM), the total energy is

$$E_{\text{tot}}(\boldsymbol{x}) = U_{\text{tot}}(\boldsymbol{x}) - \sum_{i=1}^{N} \bar{\boldsymbol{F}}_i \cdot \boldsymbol{x}_i \tag{1}$$

where $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N)^T$ represents the position of all atoms, and $\bar{\boldsymbol{F}}_i$ is the external force (if there is any) exerted on atom i. The state of minimal energy corresponds to

$$\frac{\partial E_{\text{tot}}}{\partial \boldsymbol{x}} = 0 \tag{2}$$

The Taylor expansion of E_{tot} around an initial guess $\boldsymbol{x}^{(0)} = (\boldsymbol{x}_1^{(0)}, \boldsymbol{x}_2^{(0)}, \ldots, \boldsymbol{x}_N^{(0)})^T$ of the equilibrium state gives

$$\begin{aligned} E_{\text{tot}}(\boldsymbol{x}) \approx\ & E_{\text{tot}}(\boldsymbol{x}^{(0)}) \\ & + \left.\frac{\partial E_{\text{tot}}}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\boldsymbol{x}^{(0)}} \cdot (\boldsymbol{x} - \boldsymbol{x}^{(0)}) \\ & + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^{(0)})^T \cdot \left.\frac{\partial^2 E_{\text{tot}}}{\partial \boldsymbol{x} \partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\boldsymbol{x}^{(0)}} \cdot (\boldsymbol{x} - \boldsymbol{x}^{(0)}) \end{aligned} \tag{3}$$

Its substitution into Eq. (2) yields the following governing equation for the displacement $\boldsymbol{u} = \boldsymbol{x} - \boldsymbol{x}^{(0)}$,

$$\boldsymbol{K} \cdot \boldsymbol{u} = \boldsymbol{P} \tag{4}$$

where $\boldsymbol{K} = \partial^2 E_{\text{tot}}/\partial \boldsymbol{x} \partial \boldsymbol{x}|_{\boldsymbol{x}=\boldsymbol{x}^{(0)}} = \partial^2 U_{\text{tot}}/\partial \boldsymbol{x} \partial \boldsymbol{x}|_{\boldsymbol{x}=\boldsymbol{x}^{(0)}}$ is the stiffness matrix, $\boldsymbol{P} = -\partial E_{\text{tot}}/\partial \boldsymbol{x}|_{\boldsymbol{x}=\boldsymbol{x}_{(0)}} = \bar{\boldsymbol{F}} - \partial U_{\text{tot}}/\partial \boldsymbol{x}|_{\boldsymbol{x}=\boldsymbol{x}_{(0)}}$ is the non-equilibrium force vector, and $\bar{\boldsymbol{F}} = (\bar{\boldsymbol{F}}_1, \bar{\boldsymbol{F}}_2, \ldots, \bar{\boldsymbol{F}}_N)^T$. Equation (4) is identical to the governing equation in FEM. For linear systems, the minimal energy state $\boldsymbol{x} = \boldsymbol{x}^{(0)} + \boldsymbol{u}$ can be obtained by solving Eq. (4) only once. For nonlinear system, however, Eq. (4) must be solved iteratively until $\boldsymbol{P}$ reaches zero, $\boldsymbol{P} = 0$. During each iteration step, the state variables are updated via $\boldsymbol{x}_{\text{new}}^{(0)} = \boldsymbol{x}_{\text{old}}^{(0)} + \boldsymbol{u}$, $\boldsymbol{K} = \partial^2 E_{\text{tot}}/\partial \boldsymbol{x} \partial \boldsymbol{x}|_{\boldsymbol{x}=\boldsymbol{x}_{\text{new}}^{(0)}}$, and $\boldsymbol{P} = -\partial E_{\text{tot}}/\partial \boldsymbol{x}|_{\boldsymbol{x}=\boldsymbol{x}_{\text{new}}^{(0)}}$.

*Author to whom correspondence should be addressed.

Liu et al.[2, 3] constructed a new element in AFEM to account for atomistic interactions, and demonstrated that AFEM is an order-N method for systems with sparse stiffness matrix $\boldsymbol{K}$, and is much faster than the widely used conjugate gradient method which is an order-N^2 method.

For minimization problems, FEM will converge and be stable if the total energy E_{tot} decreases during every iteration step. Since the non-equilibrium force $\boldsymbol{P} = -\partial E_{\text{tot}}/\partial \boldsymbol{x}|_{\boldsymbol{x}=\boldsymbol{x}_{\text{new}}^{(0)}}$ represents the steepest descent direction of E_{tot}, the stability and convergence are ensured if

$$\boldsymbol{u} \cdot \boldsymbol{P} > 0 \tag{5}$$

where $\boldsymbol{u}$ is the displacement increment. A sufficient condition for Eq. (5) is that the stiffness matrix $\boldsymbol{K}$ is positive definite. For problems involving material softening or nonlinear bifurcation, the stiffness matrixes $\boldsymbol{K}$ is non-positive definite, and some measures should be taken to ensure the stability condition in Eq. (5). We develop several algorithms in this paper to stabilize the FEM simulation for energy minimization problems. These are discussed in Sections 2 and 3 for direct linear equation solvers and for iterative solvers, respectively.

2. STABLE ALGORITHM FOR THE DIRECT LINEAR EQUATION SOLVER IN FEM

Examples of the direct linear equation solver include the Gaussian elimination or LU decomposition methods.

2.1. Single Degree of Freedom

We first study the simplest minimization problem—the system with a single degree of freedom. Figure 1 shows a schematic diagram of the total energy E_{tot} as a function of position x, and the point A corresponds to the energy minimum. If the point B is taken as the initial guess $x^{(0)} = x_B$, it is clear that $P_B = -\partial E_{\text{tot}}/\partial x|_{x=x_B} < 0$ and $K_B = \partial^2 E_{\text{tot}}/\partial x^2|_{x=x_B} > 0$ such that $u = x - x_B = P_B/K_B < 0$,

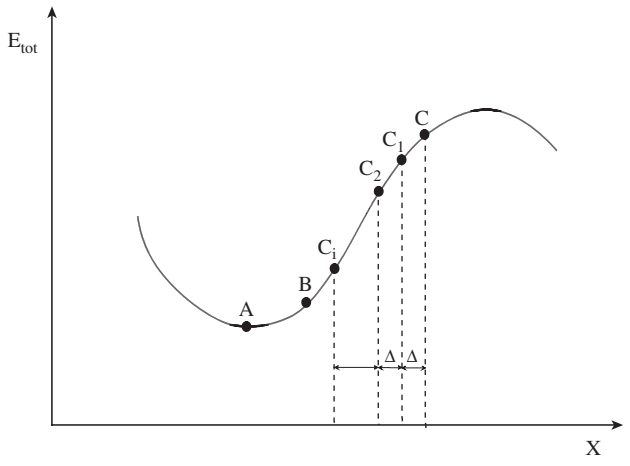

Fig. 1. A schematic diagram for one-dimensional minimization.

which moves towards the minimum state A. By repeating this process the energy minimum is reached quickly. However, when point C is taken as the initial guess, $P_C = -\partial E_{\text{tot}}/\partial x|_{x=x_C} < 0$ and $K_C = \partial^2 E_{\text{tot}}/\partial x^2|_{x=x_C} < 0$ such that $u = x - x_C = P_C/K_C > 0$, it moves in the opposite direction towards the energy minimum. Therefore, for $K = \partial^2 E_{\text{tot}}/\partial x^2 \leq 0$, we use only the first order derivative of the total energy, $P = -\partial E_{\text{tot}}/\partial x$, to search for the minimal energy. We write the displacement as

$$u = \text{sign}(P)\Delta \tag{6}$$

where sign(.) is the sign function, namely,

$$\text{sign}(P) = \begin{cases} 1, & P > 0 \\ 0, & P = 0 \\ -1, & P < 0 \end{cases} \tag{7}$$

and Δ is a positive constant. The displacement can then be written as

$$u = \begin{cases} P/K, & \text{for } K > 0 \\ \text{sign}(P)\Delta, & \text{for } K \leq 0 \end{cases} \tag{8}$$

Figure 1 illustrates the above approach, and the constant displacement increment Δ is marked in the figure. For the initial guess C at which $K_C = \partial^2 E_{\text{tot}}/\partial x^2|_{x=x_C} < 0$ and $P_C = -\partial E_{\text{tot}}/\partial x|_{x=x_C} < 0$, the resulting point after the first iteration is $x_{C_1} = x_C - \Delta$, which serves as the initial guess in the next iteration, and the corresponding K_{C_1} and P_{C_1} can be computed. Since $K_{C_1} < 0$ and $P_{C_1} < 0$ (Fig. 1), the next resulting point is $x_{C_2} = x_{C_1} - \Delta$. The above iteration process is repeated until $K_{C_i} > 0$, and then Eq. (4) is used to determine the next u. It should be pointed out that the magnitude of constant displacement increment Δ has a strong influence on the speed of convergence. Small Δ leads to a large number of iteration steps in the region with negative $K = \partial^2 E_{\text{tot}}/\partial x^2$, but large Δ may miss the minimal energy.

Equation (8) provides an efficient and robust way to find the energy minimum for one-degree-of-freedom system. The systems with multiple degrees of freedom are discussed in the following.

2.2. Multiple Degrees of Freedom

For the system with multiple degrees of freedom, the Taylor expansion of the total energy becomes

$$E_{\text{tot}} = E_{\text{tot}}^{(0)} - \boldsymbol{P}^T \cdot \boldsymbol{u} + \frac{1}{2}\boldsymbol{u}^T \cdot \boldsymbol{K} \cdot \boldsymbol{u} \tag{9}$$

where $E_{\text{tot}}^{(0)} = E_{\text{tot}}(\boldsymbol{x}^{(0)})$, and $\boldsymbol{u} = \boldsymbol{x} - \boldsymbol{x}^{(0)}$, $\boldsymbol{P} = -\partial E_{\text{tot}}/\partial \boldsymbol{x}|_{\boldsymbol{x}=\boldsymbol{x}^{(0)}}$, and $\boldsymbol{K} = \partial^2 E_{\text{tot}}/\partial \boldsymbol{x}\partial \boldsymbol{x}|_{\boldsymbol{x}=\boldsymbol{x}^{(0)}}$ which is real and symmetric, and can be expressed via the LU decomposition as

$$\boldsymbol{K} = \boldsymbol{L} \cdot \boldsymbol{U} = \boldsymbol{L} \cdot \boldsymbol{D} \cdot \boldsymbol{L}^T \tag{10}$$

where $\boldsymbol{L}$ is a lower triangular matrix, $\boldsymbol{U}$ is an upper triangular matrix, and $\boldsymbol{D}$ is a diagonal matrix. Without losing generality, we assume the diagonal components of $\boldsymbol{L}$ are 1. The substitution of Eq. (10) into Eq. (9) yields

$$E_{\text{tot}} = E_{\text{tot}}^{(0)} - \boldsymbol{P}^T \cdot \boldsymbol{u} + \frac{1}{2}\left(\boldsymbol{u}^T \cdot \boldsymbol{L}\right) \cdot \boldsymbol{D} \cdot \left(\boldsymbol{L}^T \cdot \boldsymbol{u}\right) \tag{11}$$

Let

$$\tilde{\boldsymbol{u}} \equiv \boldsymbol{L}^T \cdot \boldsymbol{u} \tag{12}$$

and Eq. (11) becomes

$$\begin{aligned} E_{\text{tot}} &= E_{\text{tot}}^{(0)} - \left(\boldsymbol{L}^{-1} \cdot \boldsymbol{P}\right)^T \cdot \tilde{\boldsymbol{u}} + \frac{1}{2}\tilde{\boldsymbol{u}}^T \cdot \boldsymbol{D} \cdot \tilde{\boldsymbol{u}} \\ &= E_{\text{tot}}^{(0)} - \tilde{\boldsymbol{P}}^T \cdot \tilde{\boldsymbol{u}} + \frac{1}{2}\tilde{\boldsymbol{u}}^T \cdot \boldsymbol{D} \cdot \tilde{\boldsymbol{u}} \end{aligned} \tag{13}$$

where

$$\tilde{\boldsymbol{P}} \equiv \boldsymbol{L}^{-1} \cdot \boldsymbol{P} \tag{14}$$

Since $\boldsymbol{D}$ is a diagonal matrix, the energy in Eq. (13) becomes

$$E_{\text{tot}} = E_{\text{tot}}^{(0)} + \sum\nolimits_i \left(-\tilde{u}_i \tilde{P}_i + \frac{1}{2} D_{ii} \tilde{u}_i^2\right) \tag{15}$$

i.e., the total energy can be decomposed into the sum of energy associated with each degree of freedom, and the approach in Section 2.1 for one-degree-of-freedom system can be adopted. We use $\tilde{\boldsymbol{u}}^+$ and $\tilde{\boldsymbol{u}}^-$ to denote $\tilde{\boldsymbol{u}}$ for the positive and negative D_{ii}, respectively, and they are given by

$$\tilde{u}_i^+ = \begin{cases} \dfrac{\tilde{P}_i}{D_{ii}}, & \text{if } D_{ii} > 0 \\ 0, & \text{if } D_{ii} \le 0 \end{cases} \tag{16a}$$

$$\tilde{u}_i^- = \begin{cases} 0, & \text{if } D_{ii} > 0 \\ \text{sign}(\tilde{P}_i), & \text{if } D_{ii} \le 0 \end{cases} \tag{16b}$$

Equation (12) then gives the corresponding parts for the displacement as

$$\boldsymbol{u}^+ = \boldsymbol{L}^{-T} \cdot \tilde{\boldsymbol{u}}^+ \tag{17a}$$

$$\boldsymbol{u}^- = \boldsymbol{L}^{-T} \cdot \tilde{\boldsymbol{u}}^- \tag{17b}$$

The displacement in this iteration step can be written as the linear superposition of these two parts as

$$\boldsymbol{u} = \alpha \boldsymbol{u}^+ + \beta \boldsymbol{u}^- \tag{18}$$

where α and β are positive numbers given in the following. It is usually required that the increment of the equivalent strain $\varepsilon = \varepsilon(\boldsymbol{u})$ (or any other scalar measurement of strain) be less than a critical value $\varepsilon_{\text{limit}}$, i.e.,

$$\varepsilon \le \varepsilon_{\text{limit}} \tag{19}$$

In order to satisfy the above constraint, α and β can be taken as

$$\alpha = \min\left(1, \frac{\varepsilon_{\text{limit}}}{2\varepsilon(\boldsymbol{u}^+)}\right) \tag{20a}$$

$$\beta = \min\left(1, \frac{\varepsilon_{\text{limit}}}{2\varepsilon(\boldsymbol{u}^-)}\right) \tag{20b}$$

such that the strain due to $\alpha \boldsymbol{u}^+$ is less than $\varepsilon_{\text{limit}}/2$, and so is the strain due to $\beta \boldsymbol{u}^-$. If all D_{ii} are positive (i.e., positive definite stiffness matrix $\boldsymbol{K}$), β is one and $\beta \tilde{\boldsymbol{u}}^-$ vanishes since $\tilde{\boldsymbol{u}}^-$ in Eq. (16b) and $\boldsymbol{u}^-$ in Eq. (17b) vanish; $\boldsymbol{u}^+$ is identical to that obtained from Eq. (4) for direct linear equations solver; and $\boldsymbol{u} = \boldsymbol{K}^{-1} \cdot \boldsymbol{P}$ represents the fastest global searching direction for energy minimum. However, for any $D_{ii} \le 0$ (i.e., non-positive definite stiffness matrix $\boldsymbol{K}$), $\boldsymbol{u} = \boldsymbol{K}^{-1} \cdot \boldsymbol{P}$ may give a slow or even wrong searching direction, but Eqs. (17) and (18) still give the displacement that is along the fastest global searching direction for energy minimum.

3. STABLE ALGORITHM FOR THE ITERATIVE LINEAR EQUATION SOLVER IN FEM

It is sometimes advantageous to use an iterative solver [e.g., $\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x})$, or $\boldsymbol{x}_{n+1} = \boldsymbol{f}(\boldsymbol{x}_n)$] in the finite element if the stiffness matrix $\boldsymbol{K}$ is positive definite. For a non-positive definite $\boldsymbol{K}$, we replace it with a modified, positive definite stiffness matrix

$$\boldsymbol{K}^* = \boldsymbol{K} + \lambda \boldsymbol{B} \tag{21}$$

where $\boldsymbol{B}$ is a positive definite matrix, and λ is a positive number to ensure the positive definiteness of $\boldsymbol{K}^*$. The resulting governing equation then becomes

$$\boldsymbol{K}^* \cdot \boldsymbol{u} = \boldsymbol{P} \tag{22}$$

It is important to point out that the state of minimal energy is independent of λ because the energy minimum is characterized by $\boldsymbol{P} = 0$. At (or near) the state of minimum energy, the stiffness matrix $\boldsymbol{K}$ is positive definite such that the modification of $\boldsymbol{K}$ to $\boldsymbol{K}^*$ is unnecessary and λ becomes zero. On the other hand, λ should not be too large such that $\boldsymbol{K}$ is not overwhelmed by $\lambda \boldsymbol{B}$ in Eq. (21). We suggest two choices for the positive definite matrix $\boldsymbol{B}$. One is the identity matrix $\boldsymbol{I}$, and the other is the mass matrix $\boldsymbol{M}$.

For an iterative solver, the positive definiteness of stiffness matrix $\boldsymbol{K}$ (or $\boldsymbol{K}^*$) can be determined from its smallest eigenvalue. If the smallest eigenvalue is positive, $\boldsymbol{K}$ (or $\boldsymbol{K}^*$) is positive definite. For a sparse $\boldsymbol{K}$ (or $\boldsymbol{K}^*$), the computation of the smallest eigenvalue is straightforward. In fact, many commercial finite element programs (e.g., ABAQUS) inform the users whether the system has negative eigenvalues and therefore whether the stiffness matrix is non-positive definite.

4. NUMERICAL EXAMPLE

We use the example of a 6 nm-long, (7,7) armchair carbon nanotube under compression to illustrate the stable

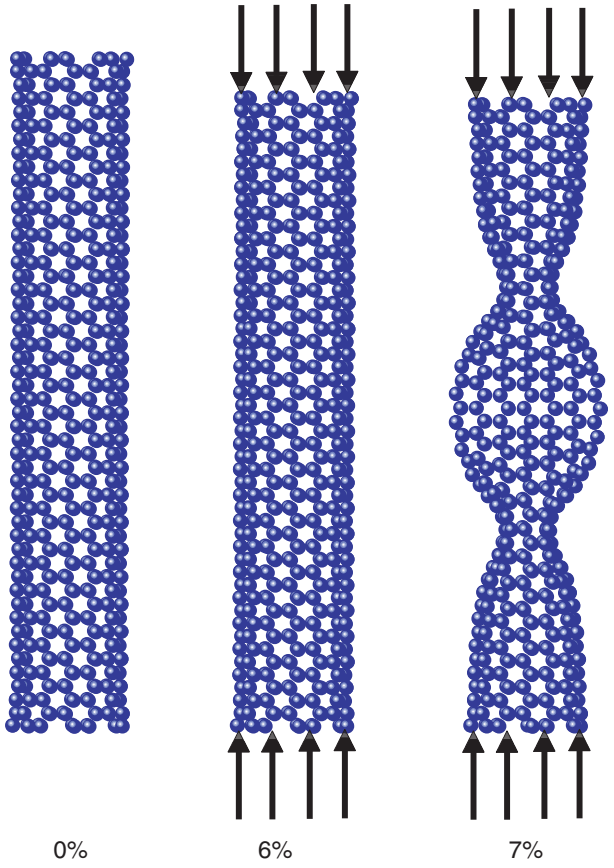

Fig. 2. Deformation patterns for a 6 nm-long (7,7) carbon nanotube under compression. Two stable algorithms proposed in Sections 2 and 3 are used in the atomic-scale finite element method. Bifurcation occurs at 7% compressive strain.

algorithms proposed above. The atomic-scale finite element method[2, 3] is used, in which each element consists of ten carbon atoms (one master atom, three nearest-neighbor atoms, and six second nearest-neighbor atoms) to account for the multi-body atomistic interactions represented by the interatomic potentials for carbon.[4, 5] These interatomic potentials display softening behavior such that the resulting stiffness matrix $\boldsymbol{K}$ may become non-positive definite. The non-positive definite $\boldsymbol{K}$ may also appear when the system is near the state of bifurcation. Figure 2 shows three stages of a carbon nanotube at the compression strain of 0%, 6% (prior to buckling), and 7% (post-buckling). The buckling pattern and the corresponding buckling strain (7%) agree well with Yakobson et at.'s molecular mechanics studies.[6] The stiffness matrix $\boldsymbol{K}$ experiences non-positive definiteness between the last two stages, but becomes positive definite again near the final stage. The proposed two algorithms in Sections 2 and 3 become necessary for such a problem involving non-positive definite stiffness matrix, and they give the same results shown in Figure 2.

It should be pointed out that, for this problem, the algorithm for the direct solver is more efficient (i.e., less incremental steps) than the algorithm for the iterative solver. This is because the algorithm for the direct solver uses the stiffness matrix $\boldsymbol{K}$ in minimization, and $\boldsymbol{K}$ is the second-order derivative of energy and governs the direction towards energy minimum. The algorithm for the iterative solver, however, uses the modified stiffness matrix $\boldsymbol{K}^*$, which is not the second-order derivative of energy anymore.

5. CONCLUDING REMARKS

We have developed two stable algorithms for the direct linear equation solver and iterative solver with non-positive definite stiffness matrix. For a direct solver with non-positive definite stiffness matrix, energy minimization of a system with multiple degrees of freedom (DOF) is decomposed to the minimization of many 1-DOF systems, and for the latter an efficient and robust minimization method is developed to ensure that the system energy decreases in every incremental step, regardless of the positive definiteness of the stiffness matrix. For an iterative solver, the stiffness matrix is modified to ensure the convergence, and the modified stiffness matrix indeed leads to the correct solution.

Acknowledgment: BL acknowledges the support of the National Natural Science Foundation of China through the Grant #10542001, #10702034, #10732050. Y. Huang acknowledges the support from NSF.

References

1. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, Numerical Recipes, Cambridge Univ. Press, New York **(1986)**.
2. B. Liu, Y. Huang, H. Jiang, S. Qu, and K.-C. Hwang, *Computer Methods in Applied Mechanics and Engineering* 193, 1849 **(2004)**.
3. B. Liu, H. Jiang, Y. Huang, S. Qu, M.-F. Yu, and K.-C. Hwang, *Phys. Rev. B* 72, 035435 **(2005)**.
4. D. W. Brenner, *Phys. Rev. B* 42, 9458 **(1990)**.
5. D. W. Brenner, O. A. Shenderova, J. A. Harrison, S. J. Stuart, B. Ni, and S. B. Sinnott, *J. Phys.-Condens. Matter* 14, 783 **(2002)**.
6. B. I. Yakobson, C. J. Brabec, and J. Bernholc, *Phys. Rev. Lett.* 76, 2511 **(1996)**.

Received: 16 August 2007. Accepted: 20 September 2007.