Searching for Alloy Configurations with Target Physical Properties: Impurity Design via a Genetic Algorithm Inverse Band Structure Approach

S. V. Dudiy and Alex Zunger*

National Renewable Energy Laboratory, Golden, Colorado 80401, USA

(Received 7 September 2005; published 25 July 2006)

The ability to artificially grow different configurations of semiconductor alloys—random structures, spontaneously ordered and layered superlattices—raises the issue of how different alloy configurations may lead to new and different alloy physical properties. We address this question in the context of nitrogen impurities in GaP, which form deep levels in the gap whose energy and optical absorption sensitively depend on configuration. We use the ''inverse band structure'' approach in which we first specify a desired target physical property (such as the deepest nitrogen level, or lowest strain configuration), and then we search, via genetic algorithm, for the alloy atomic configurations that have this property. We discover the essential structural motifs leading to such target properties. This strategy opens the way to efficient alloy design.

DOI: 10.1103/PhysRevLett.97.046401

trarily chosen set of cluster configurations will miss many nontrivial cases of N clusters with interesting physical properties. In this respect, it appears desirable to pose the ''inverse-design'' question [10]: Instead of asking which physical properties are manifested by given atomic configurations, we can ask which configuration(s) have particularly interesting given physical properties. In particular, we ask which configurations of substitutional N_p clusters have the deepest or the shallowest impurity energy level within the GaP band gap, or the highest or lowest

target value. We execute the GA search via the IAGA code [19], which utilizes the PGApack library [20]. Further implementation details can be found in Ref. [21]. In all our calculations, we start with an initial population of $N_{\rm pop} = 150$ randomly generated structures (genomes), and then replace $N_{\rm rep} = 31$ least fit genomes at each generation, using uniform crossover rate and mutation probability [18,21] of 0.25 and 0.03, respectively. We allow each GA run to continue for 40 generation (1360 property evaluations in total). At so many generations our population typically finishes its most explorative phase and does some refining of the fittest structures [22]. With the chosen GA settings, in a search for minimum VFF strain energy of a N_4 cluster, 7 out of 15 independent test runs find the configuration with the absolute minimum of the strain energy. Note that if instead of a GA run we would take a set of 1360 configuration chosen at random we would have only $\sim$0.06 probability to capture that absolute minimum strain energy configuration. We used those GA settings to search for maximum and minimum value for each of the three properties we study, and for each N_p cluster order $p = 2$–5, performing one GA run for each of those 24 cases.

The final configurations exhibiting our target properties for most of the considered target cases are depicted in Fig. 2. While this figure conveys the full configurational elements (pairs, many-body motifs) of each structure, it is interesting to observe what type of nitrogen-nitrogen pairs lead to a given target property. This is summarized in Table I. We number there only the anion fcc sublattice and denote by nn1, nn2, . . ., nn13 the first, second, and 13th nearest-neighbor (nn) nitrogen pairs occuring within clusters N_p of p nitrogen atoms. We see, for example, that if we have only two nitrogen atoms (N_2) then the minimum strain is achieved by a second nearest-neighbor (nn2) arrangement, whereas the maximum oscillator strength is achieved when the two nitrogens are 13th nearest neighbor (nn13) to each other.

Figure 2 and Table I reveal the main pair motifs responsible for given target properties. We see that: (i) Minimum strain requires the nn2 motif, but maximum strain requires the nn1 motif. (ii) Minimum VBM-to-CS transition energy (E_g) requires the nn4 motif, whereas maximum transition energy requires the nn3 one (except for the N_2 case).

esting physics or chemistry, the sheer identification of these and other recurring motifs (inspect Fig. 2 and Table I) out of an astronomic number of possibilities may provide in the future the necessary physical intuition needed for establishing ''design'' rules of materials with such target properties.

Having identified the critical areas of configuration space where our system properties are extremal, we are now in a position to learn about the overall structure-property relationship in our systems. In particular, we can take all the structure-property information we have accumulated near various extremal physical properties and ''interpolate'' such information to the rest of the configuration space, between those extremes, using a data mining approach. Let us take, as an example, the optical transition energy E_g for N_p clusters. As discussed above (viz. Table I), the configurations with extremal transition energy (Fig. 2) suggest possible correlation between the optical gap and the number of various nearest-neighbor pairs occuring within an N_p cluster. To investigate such correlations we apply multiple linear regression technique [23], using as predictor variables the numbers of occurrences m_q of various qth nearest-neighbor pairs (nn$q$ motifs) within a given N_p cluster. In particular, we model E_g as a linear function of structural parameters m_q of the cluster, as $E_g \simeq \sum_q a_q m_q$. We obtain the regression coefficients a_q using the method of least squares to minimize the error between the regression estimates and the actual data. For example, for $p = 4$ and $q = 1..6$, we get the following expression for the optical gap (in units of meV) as a function of the N_4 cluster configuration:

$$E_g \simeq 223 - 30m_1 + 7m_2 + 10m_3 - 66m_4 + 5m_5 + 0.3m_6. \quad (1)$$

The overall accuracy of this expression is described by Fig. 3. Equation (1) is seen to capture the overall trends in the property-structure relationship in this case. Moreover, from the coefficients of that equation we can gauge the relative contribution of different N-N separations within the cluster.

In conclusion, we have shown how the IBS approach [10] can be applied to efficiently explore the structure-energy relationships in a complex system, taking as an example the nontrivial case of N_p impurity clusters in GaP. The present strategy can be used to efficiently investigate such relationships in other complex systems, such as those encountered in alloy design.

This work is supported by the U.S. Department of Energy, SC-BES-DMS Contract No. DEAC36-98-GO10337.

*Electronic address: alex_zunger@nrel.gov

[1] Landolt-Börnstein, Group III Condensed Matter, Vol. 34 (Springer-Verlag, Berlin, 2001).
[2] A. Zunger and S. Mahajan, in *Handbook on Semiconductors*, edited by S. Mahajan (Elsevier Science, Amsterdam, 1994), Vol. 3, p. 1399.
[3] D. M. Eigler and E. K. Schweizer, Nature (London) **344**, 524 (1990).
[4] D. G. Thomas, J. J. Hopfield, and C. J. Frosch, Phys. Rev. Lett. **15**, 857 (1965).
[5] J. N. Baillargeon *et al.*, Appl. Phys. Lett. **60**, 2540 (1992); S. Miyoshi *et al.*, Appl. Phys. Lett. **63**, 3506 (1993).
[6] Y. Zhang *et al.*, Phys. Rev. B **62**, 4493 (2000); I. A. Buyanova *et al.*, Appl. Phys. Lett. **80**, 1740 (2002).
[7] P. R. C. Kent and A. Zunger, Appl. Phys. Lett. **79**, 2339 (2001); Phys. Rev. B **64**, 115208 (2001).
[8] L. Bellaiche, S.-H. Wei, and A. Zunger, Phys. Rev. B **54**, 17 568 (1996); **56**, 10 233 (1997); T. Mattila, L.-W. Wang, and A. Zunger, Phys. Rev. B **59**, 15 270 (1999).
[9] J. L. O'Brien *et al.*, Phys. Rev. B **64**, 161401 (2001).
[10] A. Franceschetti and A. Zunger, Nature (London) **402**, 60 (1999).
[11] J. H. Werner, S. Kolodinski, and H. J. Queisser, Phys. Rev. Lett. **72**, 3851 (1994).
[12] T. Wang *et al.*, Phys. Rev. Lett. **82**, 3304 (1999).
[13] G. H. Johannesson *et al.*, Phys. Rev. Lett. **88**, 255506 (2002).
[14] S. Kirkpatrick *et al.*, Science **220**, 671 (1983).
[15] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison-Wesley, Reading, MA, 1989).
[16] P. Keating, Phys. Rev. **145**, 637 (1966).
[17] L.-W. Wang and A. Zunger, J. Chem. Phys. **100**, 2394 (1994).
[18] D. Levine, *User's Guide to the PGAPack Parallel Genetic Algorithm Library*, Argonne National Laboratory, Argonne, Illinois Tech Report No. ANL-95/18, 1996.
[19] K. Kim and W. B. Jones, computer code IAGA, 2003.
[20] D. Levine, *et al.*, Science **220**, 671 (1983). (1998braryOp320.9(/9 s
,:5(545.49Co TDpts5(542)-257.9)-628.21.0 TD46 49E-PRL W