ELECTRONIC SUPPLEMENTARY INFORMATION

Computational Screening for Effective Ge$_{1-x}$Si$_x$ Nanowire Photocatalyst

Teck L. Tan[*] *and Man-Fai Ng*[‡]

Institute of High Performance Computing, Agency for Science, Technology and Research, 1

Fusionopolis Way, #16-16 Connexis, Singapore 138632, Singapore

[*]Address correspondence to tantl@ihpc.a-star.edu.sg
[‡]Address correspondence to ngmf@ihpc.a-star.edu.sg

Details for Density Functional Theory (DFT) Calculations

Geometry optimizations are performed using DFT implemented in the Vienna Ab-initio Simulation Package (VASP)[1,2] within a Projected Augmented Wave (PAW) basis[3] and with the Perdew, Burke and Ernzerhof (PBE) functional.[4,5] Plane-wave cutoffs are set to 400 eV and all atomic coordinates and lattices are fully relaxed until the absolute value of the forces acting on each atom is less than 0.02 eV/Å. Monkhorst-Pack Sampling with $1 \times 1 \times 8$ k-point grids is used. NWs are constructed by cleaving the bulk Si along the $\langle 100 \rangle$ growth orientation. The axis of the NWs is periodic with 4 atomic layers thick in the supercell along the $\langle 100 \rangle$ direction. We use a large supercell to ensure sufficient vacuum between each NW and its periodic images (at least 11 Å). During relaxation, the supercell volume remains fixed, although its shape is allowed to change.

Band structure and density-of-states (DOS) calculations are performed on the self-consistent charge densities. $1{\times}1{\times}19$ k-point grids are used in the DOS calculations. More accurate evaluations of band characteristics are done within the Heyd-Scuseria-Ernzerhof (HSE06) hybrid functional as implemented in VASP,[6,7] using the NW structures relaxed from PBE functional calculations. Hartree-Fock screening parameter is set to 0.2. To obtain the local potential of the unit cell, the vacuum of the supercell is expanded such that adjacent NW images are located at a distance > 20 Å from each other, enabling us to obtain the values of band edges with respect to vacuum which is set at 0 eV.

We next construct effective interactions between the occupation species Ge and Si via the cluster expansion (CE) method[8–12] using the TTK[13–19] code. The optimally truncated CE reproduces well the DFT evaluated $E_f(\sigma)$ and is used in Monte Carlo (MC) simulations to identify low-energy/stable configurations over a large search space.

The Cluster Expansion (CE) Method

In the CE method, the energy of a GeSi NW configuration, σ, is expanded in terms of cluster (cl) correlation functions $\phi_{\mathrm{cl}}(\sigma) = \prod_{k\in\mathrm{cl}} \xi_k$. The CE Hamiltonian is written as:

$$E^{CE}(\sigma) = \sum_{\mathrm{cl}} V_{\mathrm{cl}}\, \phi_{\mathrm{cl}}(\sigma). \qquad (1)$$

σ is a vector of $\{\xi_1, \xi_2, \ldots, \xi_n\}$ denoting the occupation of each binary alloy NW configuration, where ξ_k takes the value of 1 (0) if site k of the NW is occupied by a Si (Ge) atom. V_{cl} is called the effective cluster interaction (ECI). For alloy systems on a fixed lattice, a properly truncated CE Hamiltonian with a finite number of ECIs will predict structural energies accurately in most cases. Because ECIs of symmetry equivalent clusters have the same value, only symmetry-distinct ECIs in the truncated CE are evaluated, requiring only a finite set of DFT configuration

energies as the learning set. A properly truncated and converged CE should reproduce well the DFT-calculated energy of each alloy configuration, i.e., $E^{CE}(\sigma) \cong E^{DFT}(\sigma)$.

Cluster Expansion for the 2 nm Nanowire

We use the cluster expansion code, TTK,[13–19] for constructing the CE and obtain the ECI, which is then used to perform a comprehensive search for stable GeSi NW configurations using Monte Carlo simulated annealing. TTK first generates a list of symmetry-unique clusters that are ranked according to a physical hierarchy; clusters involving fewer sites and shorter spatial extent are physically more important (i.e., 2-body or pair interactions are more significant than 3-body or triplet etc., and shorter range pairs are more important than longer ranged pairs). Truncated CE sets are constructed from density functional theory (DFT) energies of relaxed NW configurations via structural inversion. Clusters in a truncated CE are selected from a large pool of clusters. Only CE that are locally complete are considered.[20] A CE basis is locally complete if it obeys all completeness relations within a truncated finite set of atomic sites and this is ensured only when each cluster in the truncated CE has all its subclusters included, e.g., if a nearest neighbor (NN) triplet cluster is included, all NN pairs and single-site clusters of the triplet has to be included. The predictive capability of the truncated CE set is evaluated via the leave-one-out cross-validation (CV_1).[12] All the above features and processes are in-built in the TTK code.

The search for ground states and stable configurations for the alloy NW is conducted iteratively. An initial learning set of ~85 configurations are used to construct the CE and via Monte Carlo simulated annealing at fixed compositions, stable configurations across the entire alloy composition are predicted. Simulated annealing starts at a high temperature of 0.15 eV/k_B (k_B is the Boltzmann constant) and cooled to 0.0002 eV/k_B at a tiny temperature step of 0.0002

eV/k_B. At each temperature step, each NW lattice site is sampled 1000 times. We note that as we are interested in the radial distribution of the atoms, the simulation box is one unit cell thick in the axial direction. The lowest energy configurations at each composition that are close to the ground state hull, if not found in the learning set, are evaluated by DFT and the energies are added to the learning set. For subsequent iterations, the CE learning sets thus contain a higher proportion of low-energy configurations. A total of 5 iterations were carried out, at which no new ground states are predicted and the structural energy difference of new stable configurations with those already in the learning set are within the error bars of DFT calculations. The final CE set contains symmetry-distinct ECI up to the third nearest neighbor pairs (multibody interactions did not improve the CV_1 score and thus not included), giving a least-squares error of 0.9 meV and a CV_1 error of 1.1 meV per atom, with the learning set (final iteration) containing 150 NW configurations. The stable configurations are shown in Fig. S1 below. The DFT formation energies of the 150 NW configurations are shown in Fig. 2 in main text.

Cluster Expansion for the 3 nm Nanowire

To obtain a CE for the 3 nm NW, we create a learning set of structures containing 130 configurations from the 2 nm NW and 9 configurations (initial guess) from the 3 nm NW. Using a unit cell that houses both the 2 nm and 3 nm NWs (widely spaced apart), TTK group symmetry-distinct ECI of both NWs using a cut-off distance setting of 4th nearest-neighbor (NN). Hence, for example, if the lattice site distribution (up to 4th NN) around a particular pair interaction in the 2 nm NW is the same as that of the 3 nm NW, these pairs are grouped under the same symmetry and assumed to have the same ECI values. The ECI is then used to search for stable alloy configurations for the 3 nm NW via Monte Carlo simulated annealing at selected

compositions. Two CE iterations were done in total and the least-squares errors for the final learning set is 1.0 meV per atom for the 2 nm NW configurations and 1.2 meV per atom for the 3 nm NW configurations. The stable configurations are shown in Fig. S2. The DFT formation energies of the 3 nm NW configurations are shown in Fig. 2 in the main text.
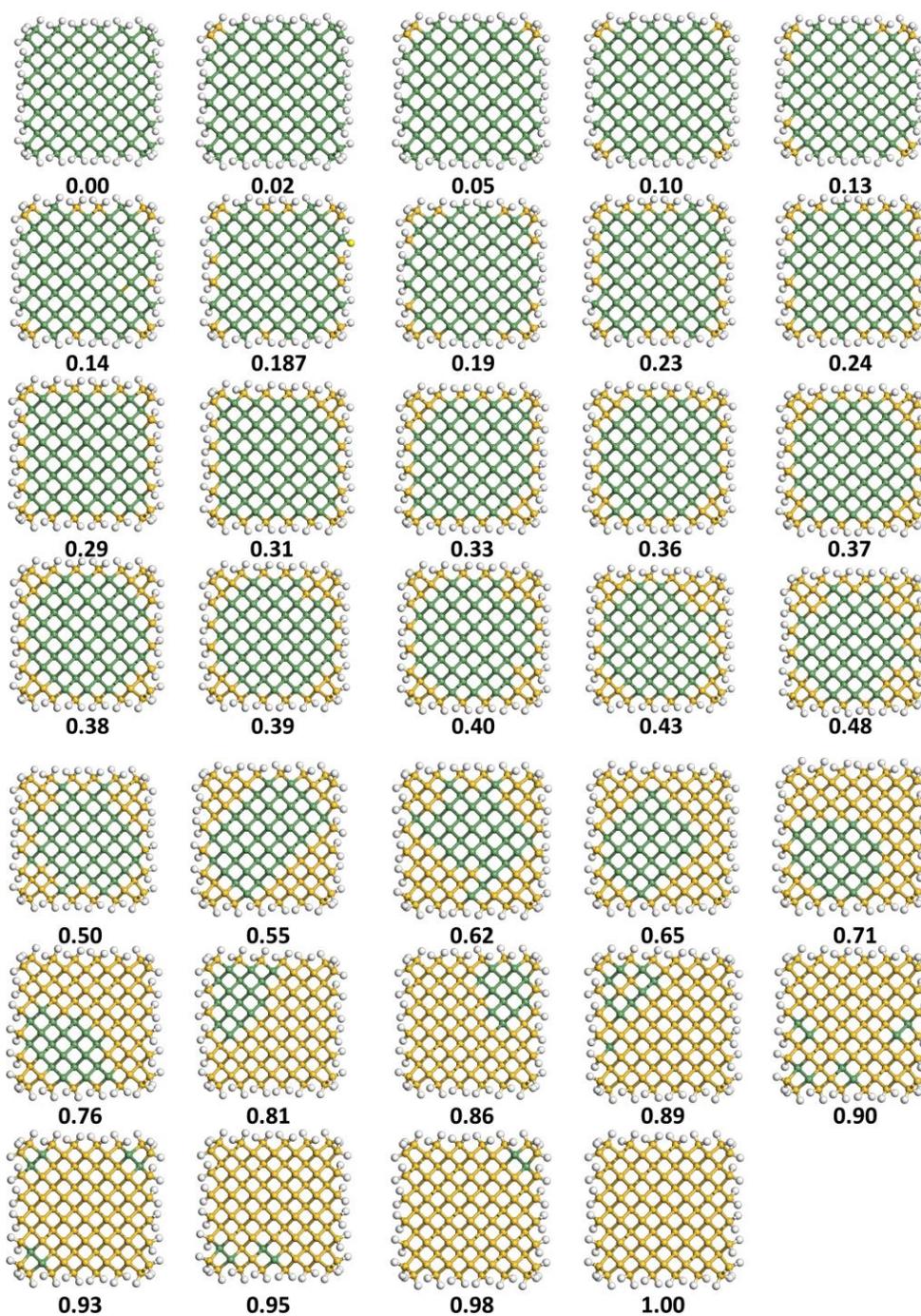
Fig. S1 Stable configurations for the 2 nm GeSi NWs across compositions. The numbers indicate the percentage of Si content (x_{Si}). Green, orange and white balls represent Ge, Si and H atoms, respectively.
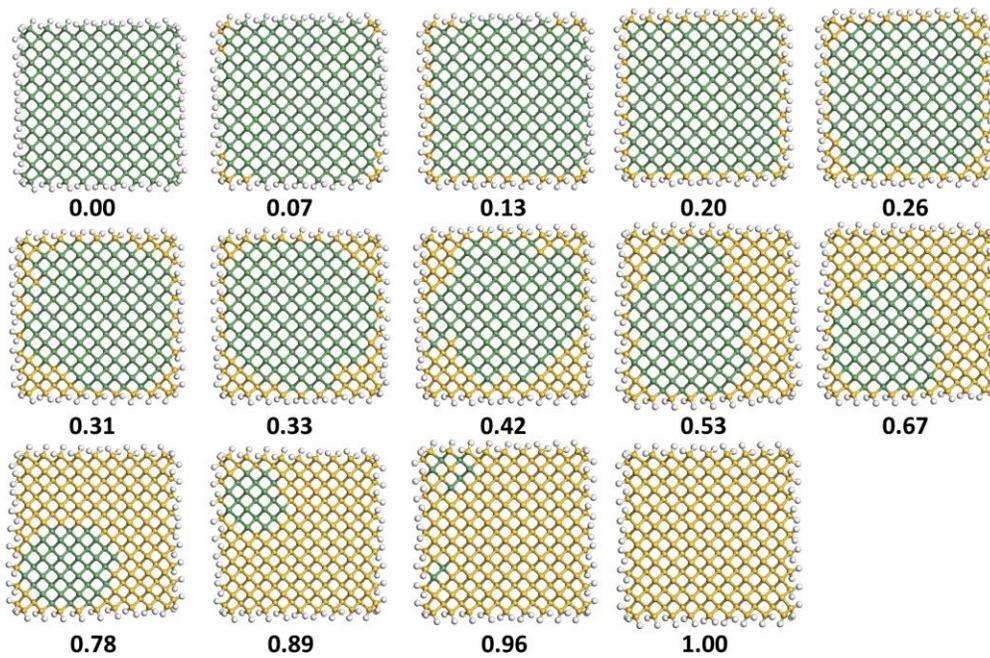
Fig. S2 Stable configurations for the 3 nm GeSi NWs across compositions. The numbers indicate the percentage of Si content (x_{Si}). Green, orange and white balls represent Ge, Si and H atoms, respectively.
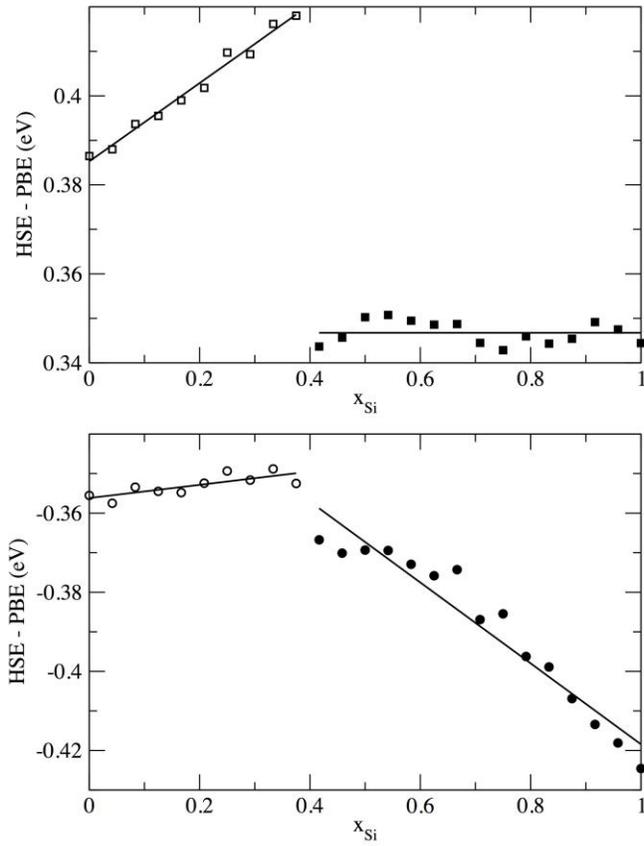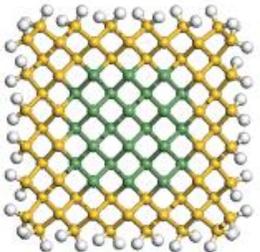
Fig. S3 Energy differences between HSE and PBE evaluated for VBM (lower panel) and CBM (upper panel) for core-shell configurations of the $\langle 100 \rangle$ GeSi NWs with diameter of 1 nm. Lines are linear fits to the points for the two different regimes $x_{Si} < 0.375$ and $x_{Si} \geq 0.375$.

Table S1 Band Edges and Bandgaps of the Symmetric and Asymmetric Core-Shell NW Structures at $x_{Si} \approx 0.71$.

Structure	x_{Si}	CBM (eV)	VBM (eV)	Bandgap (eV)	E_f (eV/atom)	Structure Illustration
Symmetric core-shell	0.71	-3.47	-4.80	1.33	-0.010	
Asymmetric core-shell	0.71	-3.60	-4.86	1.26	-0.012	
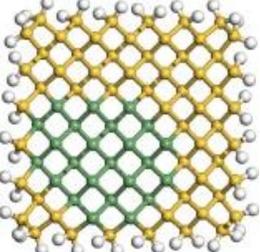
Scaling Relations Between PBE and HSE Band Edges

The valence band maximum (VBM) and conduction band minimum (CBM) for stable 1 nm NW configurations are evaluated using both PBE and HSE functionals. Their differences versus composition are shown in Fig. S3. Defining $\Delta E_{VBM} = E_{VBM}^{PBE} - E_{VBM}^{HSE}$ and likewise for CBM, we have for regime 1 ($x_{Si} < 0.375$),

$$\Delta E_{VBM} = -0.356 + 0.0167 x_{Si} \text{ and}$$

$$\Delta E_{CBM} = 0.385 + 0.088 x_{Si}.$$

And for regime 2 ($x_{Si} \geq 0.375$),

$$\Delta E_{VBM} = -0.316 - 0.102 x_{Si} \text{ and}$$

$$\Delta E_{CBM} = 0.347.$$

The values ΔE_{VBM} and ΔE_{CBM} thus varies slightly with alloy composition. Assuming the variation remains unchanged for larger NWs at fixed compositions, one could add $\Delta E_{VBM/CBM}$ to the PBE band edges of larger NW to obtain more accurate predictions. The underlying assumption is that at fixed composition, the band edge/gap shift from PBE to HSE calculated results is independent of NW size. This is analogous to previous works in demonstrating the size effect of Si NW, where the upshift between the LDA calculated bandgaps and experimental bandgaps are assumed to be the same for all NW sizes.[21–23]

Effective Cluster Interactions (ECI)

The effective cluster interactions are shown in Table S2. Here, n denotes the number of sites composing the cluster and f labels symmetry distinct clusters of the same n. For illustration, we provide a site label (corresponding to Fig. S4) of an exemplar cluster from each symmetry distinct group. For ECI with $n = 1$, a negative value indicates that the occupation of the site by Si atom is energetically favorable. For ECI with $n = 2$, a negative value indicates that the formation of Si-Si pairs is favorable. Note that for $n = 1$, the $V_{n,f}$ of shell sites are more negative than core sites, indicating that Si prefers to occupy the NW shell. The nearest-neighbor (n.n.) pair ECI are negative as well, indicating that alike atoms tend to cluster together.

Table S2 Effective Cluster Interactions (ECI or $V_{n,f}$) of the Final Cluster Expansion Set Using a Learning Set of 150 DFT NW Configurations (both 2 nm and 3 nm NWs).

n	f	Site Label	$V_{n,f}$ (eV)	Remarks
1	1	1	-0.858	core
1	2	2	-1.160	shell
1	3	3	-1.243	shell
1	4	7	-0.873	core
1	5	8	-1.203	shell
1	6	11	-0.834	core
1	7	12	-0.824	core
1	8	13	-0.854	core
1	9	19	-0.832	core
2	1	44,46	-0.170	n.n.

2	2	43,44	-0.177	n.n.
2	3	43,45	-0.009	n.n.
2	4	3,10	-0.014	n.n.
2	5	1,6	-0.046	n.n.
2	6	1,12	0.005	n.n.
2	7	7,78	-0.031	n.n.
2	8	54,56	-0.078	n.n.
2	9	13,18	-0.024	n.n.
2	10	11,20	-0.003	n.n.
2	11	2,3	0.031	2nd n.n.
2	12	66,71	0.004	2nd n.n.
2	13	64,67	0.142	2nd n.n.
2	14	64,68	-0.027	2nd n.n.
2	15	24,34	-0.006	2nd n.n.
2	16	43,52	-0.010	2nd n.n.
2	17	43,56	-0.003	2nd n.n.
2	18	1,14	-0.003	2nd n.n.
2	19	12,13	0.028	2nd n.n.
2	20	11,19	-0.014	2nd n.n.

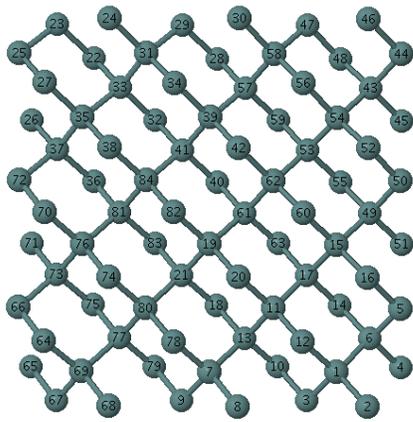**Fig. S4** Labeled lattice sites of the 2 nm NW.

REFERENCES

1   G. Kresse and J. Furthmüller, *Comput. Mater. Sci.*, 1996, **6**, 15–50.
2   G. Kresse and J. Furthmüller, *Phys. Rev. B*, 1996, **54**, 11169–11186.
3   P. E. Blöchl, *Phys. Rev. B*, 1994, **50**, 17953–17979.
4   J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
5   J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1997, **78**, 1396–1396.
6   J. Paier, M. Marsman, K. Hummer, G. Kresse, I. C. Gerber and J. G. Ángyán, *J. Chem. Phys.*, 2006, **124**, 154709.
7   J. Paier, M. Marsman, K. Hummer, G. Kresse, I. C. Gerber and J. G. Ángyán, *J. Chem. Phys.*, 2006, **125**, 249901.
8   J. M. Sanchez, F. Ducastelle and D. Gratias, *Physica*, 1984, **128A**, 334–350.
9   J. W. D. Connolly and A. R. Williams, *Phys. Rev. B*, 1983, **27**, 5169.
10   V. Blum, G. L. W. Hart, M. J. Walorski and A. Zunger, *Phys. Rev. B*, 2005, **72**, 165113.
11   V. Ozoliņš, C. Wolverton and A. Zunger, *Phys. Rev. B*, 1998, **57**, 6427.
12   A. Van de Walle and G. Ceder, *J. Phase Equilib.*, 2002, **23**, 348–359.
13   N. A. Zarkevich, T. L. Tan and D. D. Johnson, *Phys. Rev. B*, 2007, **75**, 104203.
14   N. A. Zarkevich, T. L. Tan, L.-L. Wang and D. D. Johnson, *Phys. Rev. B*, 2008, **77**, 144208.
15   L.-L. Wang, T. L. Tan and D. D. Johnson, *Phys. Rev. B*, 2012, **86**, 035438.
16   T. L. Tan, L.-L. Wang, D. D. Johnson and K. Bai, *Nano Lett.*, 2012, **12**, 4875–4880.
17   M.-F. Ng and T. L. Tan, *Nano Lett.*, 2013, **13**, 4951–4956.
18   T. L. Tan, L.-L. Wang, D. D. Johnson and K. Bai, *J. Phys. Chem. C*, 2013, **117**, 22696–22704.
19   L.-L. Wang, T. L. Tan and D. D. Johnson, *Nano Lett.*, 2014, **14**, 7077–7084.
20   N. A. Zarkevich and D. D. Johnson, *Phys. Rev. Lett.*, 2004, **92**, 255702.
21   A. J. Read, R. J. Needs, K. J. Nash, L. T. Canham, P. D. J. Calcott and A. Qteish, *Phys. Rev. Lett.*, 1992, **69**, 1232–1235.
22   B. Delley and E. F. Steigmeier, *Appl. Phys. Lett.*, 1995, **67**, 2370–2372.
23   D. D. D. Ma, C. S. Lee, F. C. K. Au, S. Y. Tong and S. T. Lee, *Science*, 2003, **299**, 1874–1877.