COMPUTATIONAL PROGRAMS FOR PHYSICS

An efficient algorithm for generating perfect special quasirandom structures with many-body correlation functions and its application to the Kronig–Penney model

COMPUTATIONAL PROGRAMS FOR PHYSICS

An efficient algorithm for generating perfect special quasirandom structures with many-body correlation functions and its application to the Kronig–Penney model

Tian-Ze Li(李天择)[1], Chang-Chun He(何长春)[1,2], and Xiao-Bao Yang(杨小宝)[1,†]

[1] *School of Physics and Optoelectronics, South China University of Technology, Guangzhou* 510640, *China*
[2] *Guangdong Provincial Key Laboratory of Functional and Intelligent Hybrid Materials and Devices, South China University of Technology, Guangzhou* 510640, *China*

(Received 15 February 2026; revised manuscript received 5 May 2026; accepted manuscript online 7 May 2026)

The special quasirandom structure (SQS) method provides an ideal representation of disordered structures. However, it is still a challenge to generate structures that perfectly satisfy the constraints of correlation functions, especially for many-body interactions. Taking one-dimensional systems as an example, we develop an efficient SQS construction algorithm for multi-component materials based on De Bruijn sequences, which can be extended to systems with higher dimensions and more components. The SQSs constructed by our algorithm are shown to universally satisfy all the constraints of $1 \sim n$-body interactions, and the numbers of consecutive identical-atom subsequences follow a unified counting rule. Based on the Kronig–Penney model, we have systematically investigated the impact of structural disorder on electronic properties. As the cell size increases, there are the same trends in the electronic structures obtained from SQSs and fully random structures, while SQSs display markedly faster convergence and significantly reduced fluctuations. We demonstrate that SQS provides a reliable and efficient finite-size representation for electronic-structure calculations of disordered systems.

Keywords: disordered structure, special quasirandom structure, De Bruijn sequence, Kronig–Penney model

PACS: 61.43.Dq, 71.23.–k, 02.70.–c

DOI: 10.1088/1674-1056/ae69bf **CSTR:** 32038.14.CPB.ae69bf

1. Introduction

The macroscopic properties of multicomponent materials are dominated by their microstructures. For instance, in the binary alloy Cu_xAu_{1-x}, a disordered structure obtained by quenching exhibits a smooth and regular dependence of the resistivity on composition, whereas the resistivity displays pronounced oscillatory variations when the atomic arrangement becomes ordered under annealing conditions.[1] In high-entropy alloys, semiconductor alloys, and oxide solid solutions, atoms are randomly arranged without periodicity, and the disappearance of translational symmetry of the Hamiltonian induces band broadening, electronic localization, and pronounced changes in transport and magnetic behaviors.[2] For disordered alloys, Voicu Popescu *et al.* proposed a spectral decomposition method based on large supercells to describe "polymorphous configurations", in which different random configurations were calculated and the spectral functions were averaged;[3] for non-crystalline structures, Matthew Jankousky *et al.* developed a first-principles-based method to generate structures with short-range order for ensemble averaging and spectral function analysis of the electronic structure, where the calculated photoelectronic properties are in good agreement with experiments.[4]

A variety of methods have been developed to describe the structural stabilities of alloys. The cluster expansion (CE) method[5] provides a systematic statistical-mechanical framework for various solid solutions by expressing the configurational energy as a linear combination of cluster correlation functions. In contrast, the similar atomic environment (SAE) approach[6] generates representative structures by minimizing the statistical deviation of atomic environments, offering an alternative strategy of structural fingerprints for the construction of machine learning potentials. For disordered alloys, the virtual crystal approximation (VCA)[7] and the coherent potential approximation (CPA)[8] provide efficient mean-field treatments by averaging the atomic properties over the random distribution of constituents. In this study, we specifically focus on the special quasi-random structure (SQS) method, which aims to construct a small periodic cell whose correlation functions are equivalent to those of an infinitely large, completely random disordered structure.[9] When combined with high-precision DFT calculations, it can be used to describe properties of disordered alloys, replacing the traditional random-occupation model and significantly improving computational efficiency.[10,11] In metallic alloys, SQS can be used to predict key properties such as mixing enthalpy, lattice constants, and magnetic moments of bcc-, fcc-, and hcp-structured

†Corresponding author. E-mail: scxbyang@scut.edu.cn

systems.[12–15] In semiconductor alloys, SQS can be used to accurately predict the electronic structure, band gap, and formation energy of III–V compounds and Si–Ge alloys.[16,17] In oxide systems, SQS can be used not only to simulate the dielectric, piezoelectric, and ferroelectric properties of perovskite oxides such as $Ba_{1-x}Sr_xTiO_3$ and $Pb(Zr_{1-x}Ti_x)O_3$, as well as the formation of polar nanoregions,[18–20] but also to investigate the oxygen vacancy distribution and ionic conductivity in doped ceria-based solid electrolytes such as $Ce_{1-x}Gd_xO_{(2-x)/2}$.[21] For ceramic and two-dimensional (2D) systems, SQS can be employed to investigate the mixing enthalpy and structural stability of MAX phases, such as Zr_2AlC-based solid solutions, providing theoretical guidance for the design of high-entropy ceramic materials.[22] In two-dimensional disordered alloys, such as $TiNbCO_2$, SQS enables reliable predictions of the formation energy, elastic properties, and electronic structure, which are valuable for the design of energy- and environment-related materials.[23] In addition, SQS can be used to analyze the effects of the random distribution of hydrogen atoms in zirconium hydrides (ZrH_{2-x}) on the mechanical properties and electronic structure of materials, further extending its applications in materials science.[24]

Despite the wide applications of SQS, there are certain limitations in existing methods for constructing SQS. For example, the simulated annealing method based on Monte Carlo[25] exhibits relatively slow convergence in searching SQS in larger structures, making it difficult to obtain an SQS whose correlation functions are perfectly satisfied. The SQS searching algorithm based on zero-suppressed binary decision diagram (ZDD)[26] considers only two-body interactions, which can solve systems containing at most 40 atoms. The small set of ordered structures (SSOS),[27] performing a weighted average over multiple small cells, only ensures that physical quantities, such as correlation functions, match their weighted averages without preserving the spatial distribution, and can efficiently simulate certain average thermodynamic properties while failing to represent the actual microscopic structure. Additionally, these SQS methods typically consider only two-body interactions due to the increased computational difficulty associated with incorporating many-body interactions. The essence of representative structures lies in matching correlation functions, and the ideal SQS should reproduce as many correlation functions as possible,[9] although most SQS constructions in practice consider only pair interactions. Recent studies have shown that in alloy systems such as zirconium, interactions beyond pair terms (e.g., three-body and four-body interactions) can be significant and cannot be ignored, and they strongly depend on the local atomic configurations.[28] Therefore, in order to accurately model realistic disordered systems, it is necessary to construct SQSs that match many-body correlation functions.

In this work, taking a one-dimensional alloy as an example, we have developed an SQS solving method constrained by the number of many-body interactions, which can efficiently obtain perfect SQS structures with various compositions. It can be proven that an SQS structure satisfying n-body interactions also satisfies all $1 \sim (n-1)$-body interactions. In the case of binary SQS, it additionally satisfies the two-body interactions of rank $1 \sim (n-1)$. For all SQS, there is also a quantitative relationship for the number of consecutive identical atoms. Combined with the Kronig–Penney model, we have performed a comparative study of the electronic structures of SQSs of different sizes and random structures. We have elucidated the physical mechanism of localization of electronic states in disordered potentials and explained the origin of their statistical behaviors by analyzing the evolution of energy, the statistical properties of energy levels, the density of states (DOS), and the spatial distribution of electronic wavefunctions with respect to supercell size. Thus, the perfect SQS obtained by our algorithm can well describe the properties of disordered structures.

2. Algorithm for constructing SQS based on many-body interactions

In order to evaluate multi-component systems, the Hamiltonian is derived based on the lattice gas model[29]

$$\begin{aligned} H_{\mathrm{LGM}} = & -\sum_i (\mu_A P_i^A + \mu_B P_i^B) \\ & - \sum_{\langle \mathrm{NN} \rangle} [J_{AA} P_i^A P_j^A \\ & + J_{AB}(P_i^A P_j^B + P_i^B P_j^A) + J_{BB} P_i^B P_j^B] + \cdots . \end{aligned} \tag{1}$$

The first term represents the on-site energy contribution, where the occupation variables $P_i^A, P_i^B \in \{0,1\}$ are coupled to their respective chemical potentials μ_A and μ_B, effectively defining the grand canonical energy of the particles. The second term accounts for the pairwise interactions between nearest neighbors, summing the bond energies J_{AA}, J_{AB}, and J_{BB} based on the configuration of occupied sites. The subsequent '$\cdots$' represents additional higher-order interaction terms.

By normalizing the coefficients, the correlation functions can be expressed in the following form:

$$\begin{aligned} \Pi_{\mathrm{LGM}} = \frac{1}{N} \Bigg(& \sum_i P_i^A, \sum_i P_i^B, \sum_{\langle \mathrm{NN} \rangle} P_i^A P_j^A, \\ & \sum_{\langle \mathrm{NN} \rangle} (P_i^A P_j^B + P_i^B P_j^A), \sum_{\langle \mathrm{NN} \rangle} P_i^B P_j^B \Bigg). \end{aligned} \tag{2}$$

In a disordered structure, the occupations of any two lattice sites are statistically independent, implying that the joint occupation probability factorizes as $\langle P_i^\alpha P_j^\beta \rangle = \langle P_i^\alpha \rangle \langle P_j^\beta \rangle = x_\alpha x_\beta$, where x_α denotes the molar fraction of component α.

Under the thermodynamic limit, the summation over nearest-neighbor (NN) pairs in Eq. (2) can be replaced by the ensemble average. Consequently, the elements of the correlation vector Π_{LGM} for an ideal random structure are determined solely by the composition

$$\Pi_{\mathrm{LGM}}^{\mathrm{rand}} = \left(x_A, x_B, \frac{z}{2}x_A^2, zx_Ax_B, \frac{z}{2}x_B^2\right), \tag{3}$$

where z is the coordination number of the lattice. For an equiatomic binary system ($x_A = x_B = 0.5$), this leads to the symmetric distribution of bond frequencies, where the total occurrence of hetero-atomic pairs (zx_Ax_B) is precisely balanced by the sum of homo-atomic pairs ($\frac{z}{2}(x_A^2 + x_B^2)$).

To simulate a random structure, one can construct a cell whose correlation function components match those of a perfectly disordered structure; such a configuration is termed an SQS.[9] Furthermore, during the construction of an SQS, more stringent constraints can be imposed: specifically, requiring the occurrences of the four types of pair clusters (*AA*, *BB*, *AB*, and *BA*) to be strictly equal. We define this condition as satisfying the pair-interaction constraint. Under this constraint, the SQS not only matches the correlation functions of the disordered state but also exhibits superior physical properties, as demonstrated in subsequent sections. While the aforementioned correlation functions only account for two-body interactions, higher-order many-body interactions are expected to follow similar statistical principles.[9] Therefore, the constraint can be further extended by requiring each type of n-body cluster configuration to appear with equal frequency, thereby satisfying the n-body interaction constraint. Generally, as the order of satisfied interaction constraints increases, a larger number of correlation function components in the SQS become equivalent to those of the ideal disordered structure.

Taking a one-dimensional binary SQS as an example, the structure can be represented as a linear sequence, where an n-body cluster corresponds to a subsequence of length n. As shown in Fig. 1, the two-body and three-body clusters are labeled, and all distinct two-body clusters and three-body clusters occur with equal frequency in the structure. This demonstrates that the structure simultaneously satisfies the statistical requirements for both two-body and three-body interactions. As n increases, more atoms can be generated in the SQS, which can satisfy higher-order many-body interactions to represent the disordered structure more accurately.

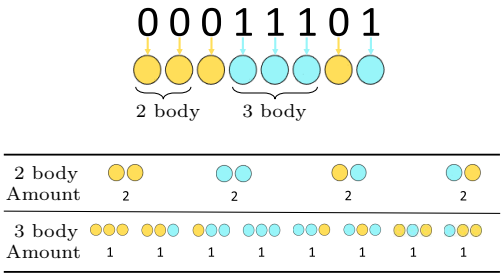

Fig. 1. Statistics of n-body cluster occurrences in a representative SQS under periodic boundary conditions.

For the construction of SQS in this work, each possible subsequence of length n appears exactly once, which is mathematically equivalent to constructing a $B(k,n)$ De Bruijn sequence.[30] A $B(k,n)$ De Bruijn sequence is a cyclic sequence over an alphabet of size k, in which every possible subsequence of length n appears exactly once. A $B(k,n)$ De Bruijn sequence exists for any positive integers k and n.[30] We now give a detailed introduction to the classical method for constructing a $B(k,n)$ De Bruijn sequence, which is based on the Eulerian cycle of the De Bruijn graph (Fig. 2).[30]

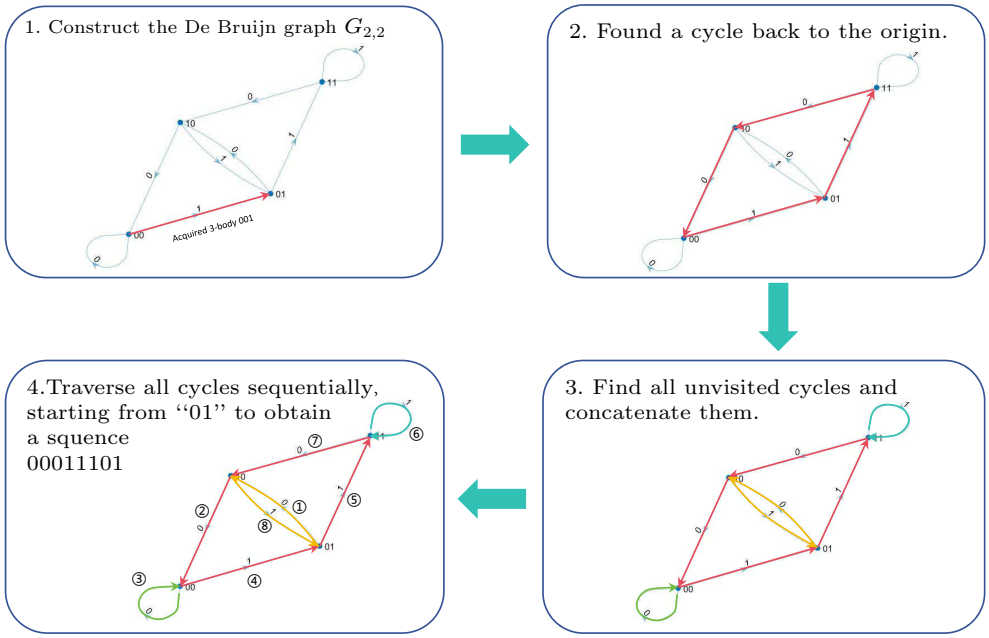

Fig. 2. A simplified flowchart of the De Bruijn sequence solving algorithm, illustrated with $B(2,3)$ as an example.

1. Constructing the De Bruijn graph $G_{k,n-1}$

Vertices: The vertex set V of the graph consists of all k-ary strings of length $n-1$, yielding a total of k^{n-1} vertices.

Edges: For any two vertices u and v, a directed edge from u to v exists if the suffix of u overlaps with the prefix of v by $n-2$ symbols. Specifically, let $u = \sigma w$ and $v = w\tau$, where w is a string of length $n-2$ and σ and τ are symbols from the alphabet. In this case, a directed edge from u to v is defined.

Edge labeling: The directed edge (u,v) corresponds to the length-n substring $\sigma w\tau$. Equivalently, the edge can be labeled by the newly appended symbol τ, indicating that it extends the preceding length-$(n-1)$ string u.

2. Finding the Eulerian cycle

An Eulerian cycle is a closed path in a graph that traverses each edge exactly once. Since the De Bruijn graph is balanced, with each vertex having in-degree and out-degree equal to k, it necessarily admits an Eulerian cycle. An Eulerian cycle can be found using algorithms such as the Hierholzer algorithm.[31]

3. Constructing the sequence

Starting from the initial vertex of the Eulerian cycle, the new symbol τ on each edge is recorded sequentially. By connecting these symbols, the $B(k,n)$ sequence is constructed.

For two-dimensional SQSs, a similar strategy can be adopted by confirming that each many-body cluster appears with the same occurrence number within the cell, which is equivalent to constructing a De Bruijn torus. The FFMS method (proposed by Fan, Fan, Ma, and Siu)[32] is a classical and efficient approach that provides a construction framework based on transforming one-dimensional De Bruijn sequences via Cartesian products; its core idea is as follows: First, a one-dimensional k-ary De Bruijn sequence is constructed. Then, multiple copies of this sequence are arranged as column vectors using a column-shift operator with a carefully designed shift vector S, such that each column is cyclically shifted relative to the previous one by a prescribed number of positions. Through this geometric offset, the local window properties of the one-dimensional sequence are effectively mapped onto a global coverage property in two dimensions, ensuring that every $r \times s$ subarray in the matrix is unique and exhaustive under periodic boundary conditions. This procedure ultimately yields a perfect two-dimensional k-ary distribution. For SQSs with different many-body clusters and different numbers of atomic species, alternative construction approaches such as the grid method[33] and the cock method[34] have also been employed.

3. Results and discussion

3.1. Generating SQS in multi-component materials

Using the above algorithm, we obtained all binary SQS solutions for $n = 1 \sim 5$, a single binary SQS solution for each $n = 1 \sim 22$, and binary SQS solutions satisfying element-exchange symmetry for $n = 3$, 5, 7, 9, and 11. In addition, single ternary SQS solutions were obtained for $n = 1 \sim 12$ (Table 1), together with partial solutions for four-, five-, and six-component SQSs. Furthermore, by applying the many-body interaction constraint, we have constructed two-dimensional SQSs that satisfy four-body and five-body constraints, respectively.

Table 1. Binary and ternary SQS sequences, where 0, 1, and 2 represent three different types of atoms. Due to the large number of atoms, only a portion of the results is shown.

n	Binary SQS sequences $B(2,n)$
2	0011
3	00011101
4	0000111101100101
5	00000111101110010100110101100001
6	000000111111011110011101011100011011010011001011000010101000 1001

n	Ternary SQS sequences $B(3,n)$
2	002212011
3	000222122021121020120011101
4	000022221222022112210220122002121202111211021012100202011201020012000111101100101

As shown in Fig. 3(a), there is an SQS with 16 sites that satisfies the 4-body interactions. Note that the numbers of 2-body and 3-body interactions are also satisfied in this structure. By examining all many-body interactions of order up to n in a given SQS, we find that, for any fixed subsequence length, all subsequences with different configurations occur with the same frequency within that SQS. This indicates that an SQS constructed to satisfy n-body interactions automatically satisfies all $1 \sim (n-1)$-body interactions as well. The counts of subsequences of different lengths in ternary SQSs are plotted on a logarithmic scale (Fig. 3(b)), and the results form straight lines. This indicates that the number of subsequences of length l increases exponentially with l, specifically following a 3^l scaling law. A detailed analysis shows that, in an n-body SQS, any l-body subsequence appears exactly k^{n-l} times. This property can be understood as follows.

In a k-component SQS satisfying n-body interactions, every n-body configuration (a total of k^n configurations) appears

exactly once. For a given l-body configuration w_l (i.e., a subsequence of length l, where $l < n$), it can be regarded as a prefix of an n-body configuration. Once this prefix is fixed, the remaining $n-l$ positions can take any values independently, resulting in k^{n-l} possible extensions. Therefore, each subsequence of length l serves as the prefix of exactly k^{n-l} different n-body configurations. Since the SQS satisfying the n-body interaction enumerates all n-body configurations without repetition or omission, every n-body configuration that starts with w_l appears exactly once. Consequently, w_l appears together with all these k^{n-l} different n-body configurations, and its occurrence number is

$$O(w_l) = k^{n-l}. \tag{4}$$

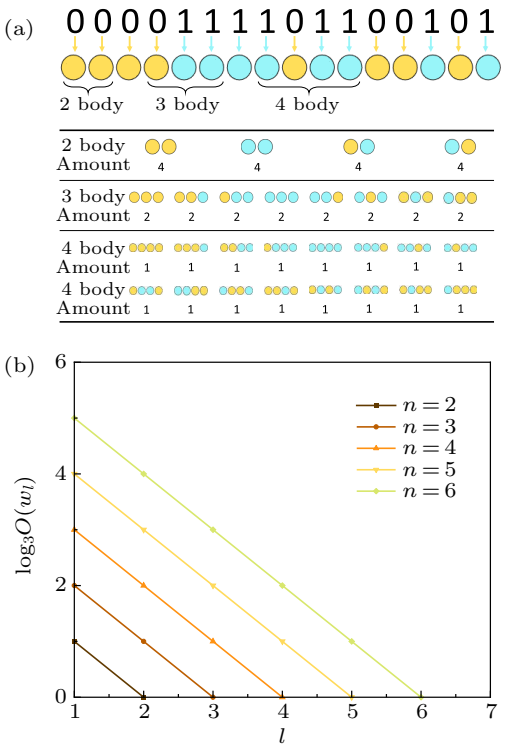

Fig. 3. (a) Schematic diagram of the binary $n = 4$ SQS, showing its 2-body to 4-body interactions, demonstrating its satisfaction of $1 \sim 4$-body interactions (the atomic number ratio ensures that the 1-body interaction is automatically satisfied). (b) Statistical analysis of the number $O(w_l)$ of ternary subsequences of length l, where the vertical axis represents $\log_3 O(w_l)$ and the horizontal axis represents the subsequence length l. Different curves correspond to SQS sequences with different values of n.

3.2. Structural characteristic in SQS

An m-th order two-body refers to a pair of atoms whose separation equals m lattice spacings. All orders of two-body interactions are illustrated in Fig. 4(a). Taking the binary SQS with $n = 3$ as an example, the numbers of first- and second-order two-body pairs are both equal to 2, indicating that this structure satisfies the first- and second-order two-body interactions. Figure 4(b) shows all orders of two-body interactions for all binary SQSs for different values of n, where there is a linear dependence between the logarithm of the number and the neighboring index i in the range $1 \le i \le n-1$. This indicates that the number of two-body pairs of orders 1 to $n-1$ is fixed and equal to 2^{n-2}. Consequently, any SQS satisfying the n-body interaction automatically satisfies the $1 \sim (n-1)$-order two-body interactions, and the corresponding $1 \sim (n-1)$ components of the two-body correlation function are exactly zero, perfectly matching those of the disordered structure. This property can be understood as follows.

For a two-body pair with a fixed separation j, it can be regarded as the first atom and the $(j+1)$-th atom in an n-body configuration, while the remaining $n-2$ sites can be freely occupied by atom A or B. Therefore, each two-body pair can be extended into 2^{n-2} different n-body configurations. In a binary SQS satisfying the n-body interaction, there are 2^n distinct n-body configurations, and each n-body configuration appears exactly once in the sequence. As a result, if a given two-body pair can be extended into 2^{n-2} different n-body configurations, it must necessarily appear exactly 2^{n-2} times in the sequence. All four types of two-body pairs, 00, 01, 10, and 11, therefore appear 2^{n-2} times, which means that the SQS satisfies the two-body interaction and perfectly matches the correlation function of the disordered structure.

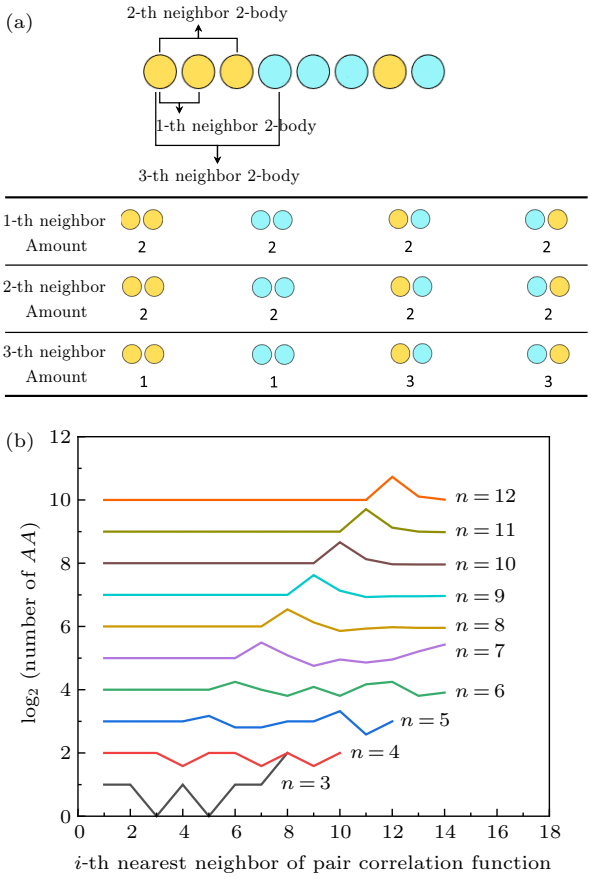

Fig. 4. (a) Schematic diagram of the binary $n = 3$ SQS, showing its one-order to three-order two-body interactions, demonstrating its satisfaction of $1 \sim 2$-order two-body interactions. (b) Statistical analysis of the two-body interaction correlation function components for SQS structures with different values of n, where the horizontal axis represents various nearest neighbors, and the vertical axis represents the number of identical atomic AA nearest neighbors at each level.

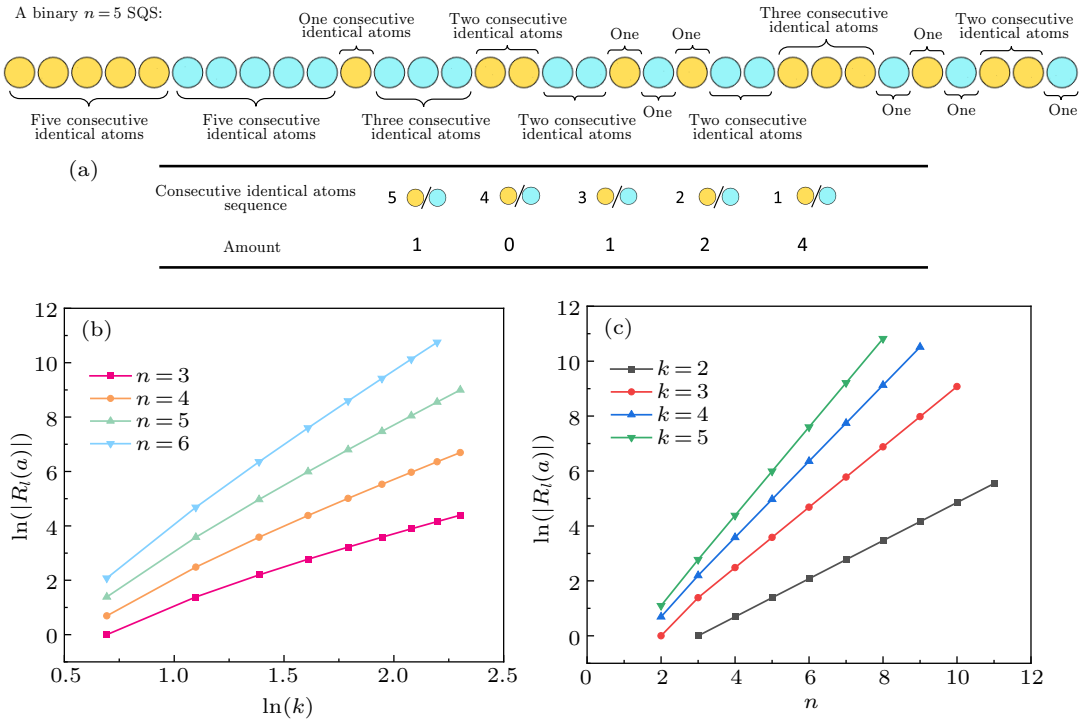

Fig. 5. (a) Schematic illustration of a binary $n=5$ SQS, showing the statistics of consecutive identical-atom sequences with different lengths. (b) The logarithm of the number of consecutive identical-atom sequences of length $\ln R_l(a)$. The vertical axis is $\ln R_l(a)$, and the horizontal axis is $\ln k$, where k denotes the number of atomic species. Different curves correspond to SQSs with different values of n. (c) The vertical axis is $\ln R_l(a)$, and the horizontal axis is n. Different curves correspond to different values of k.

By counting the numbers of consecutive identical-atom subsequences of different lengths in all SQS structures (shown in Fig. 5), we find that the data form an approximately straight line in a logarithmic plot. This indicates that the number of such subsequences follows a relation of the form number $= k^n$, depending only on k and n and being independent of the atom species. Through further analysis, the following relation can be proven:

$$|R_l| = \begin{cases} 1, & l = n, \\ k-2, & l = n-1, \\ (k-1)^2 k^{n-l-2}, & 1 \le l \le n-2, \end{cases} \tag{5}$$

where $|R_l|$ denotes the number of consecutive identical-atom subsequences of length l. This property can be understood as follows.

For $l = n$: each n-body configuration appears exactly once in an SQS; therefore, there is only one consecutive A-atom subsequence of length n.

For $l = n-1$: according to Eq. (4) in property 1, each $(n-1)$-body composed of $(n-1)$ A atoms appears exactly k times in an SQS. In the n-body A^n, A^{n-1} appears twice, but these two occurrences are not consecutive subsequences of length $n-1$. Therefore, the number of consecutive A-atom subsequences of length $n-1$ is $k-2$.

For $l \le n-2$: a consecutive A-atom subsequence of length l must take the form $xA^l y$ with $x, y \ne A$, because otherwise the length of the consecutive A block would exceed l. Since every length-n sequence appears exactly once in an SQS, fixing a length-l consecutive subsequence together with its left and right neighboring atoms and the remaining $n-l-2$ positions to the right uniquely determines a length-n sequence. For $1 \le l \le n-2$, the two boundary sites can each be chosen from $k-1$ non-A atom types, while the remaining $n-l-2$ sites can take any of the k symbols. Thus, the number of length-n sequences containing A^l as a consecutive block is

$$\underbrace{x}_{k-1 \text{ choices}} A^l \underbrace{y}_{k-1 \text{ choices}} \underbrace{z_1 z_2 \cdots z_{n-l-2}}_{k^{n-l-2} \text{ choices}},$$

which yields $(k-1)^2 k^{n-l-2}$ possible length-n sequences. Each of these sequences appears exactly once in the SQS, and hence the number of consecutive A-atom subsequences of length l is also $(k-1)^2 k^{n-l-2}$. This proves Eq. (5). Clearly, the same conclusion holds for any atom type.

3.3. Applications in the Kronig–Penney model

To demonstrate the configuration effect in the disordered structures, we employ the Kronig–Penney model to investigate the electronic properties of the generated SQSs. The Hamiltonian is given by

$$\hat{H} = -\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + V(x), \quad V(x) = \begin{cases} 0, & \text{atom } A, \\ U_0, & \text{atom } B, \\ 2U_0, & \text{atom } C, \\ \ldots, & \ldots \end{cases} \tag{6}$$

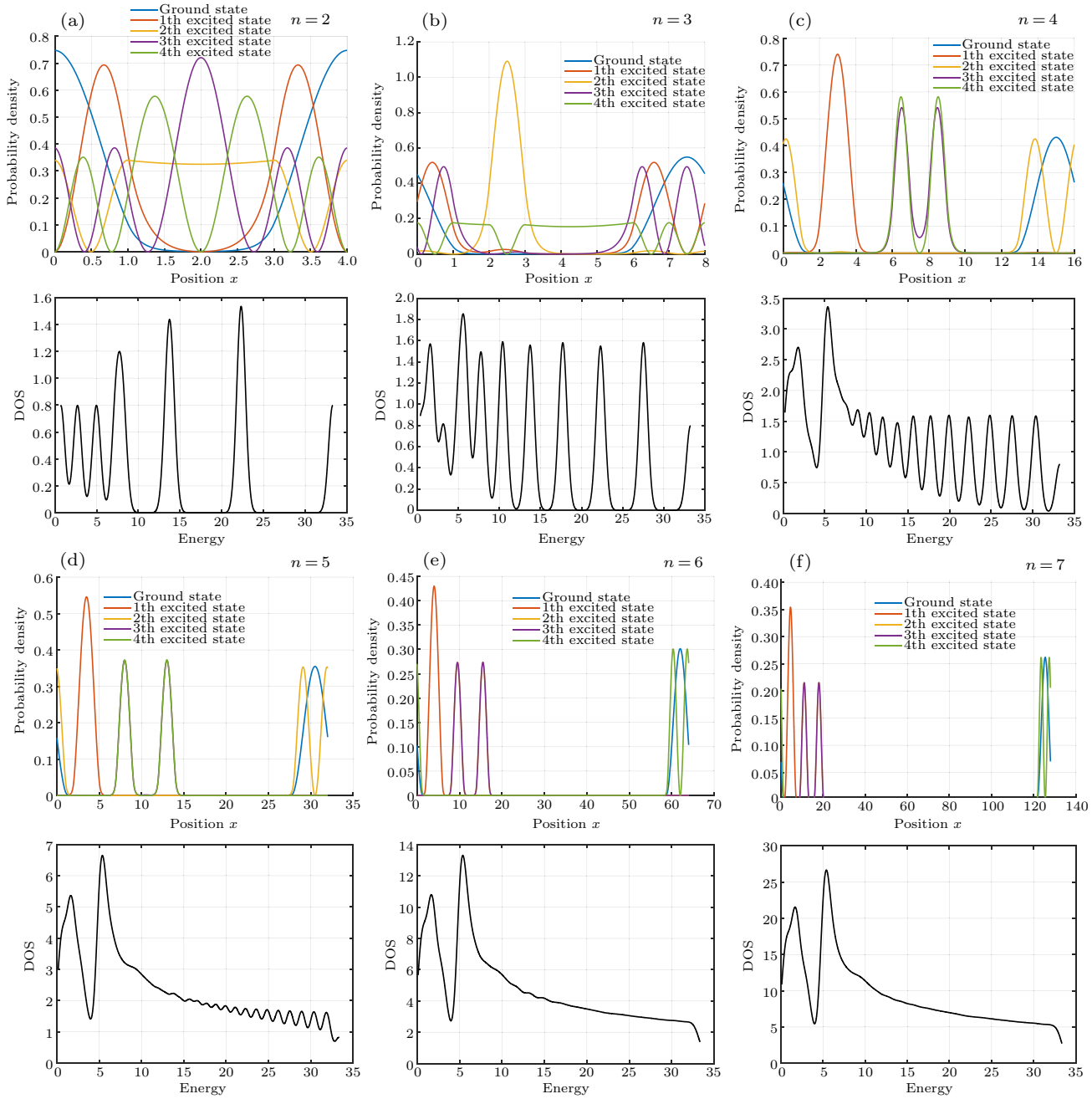

Fig. 6. (a)–(f) Wavefunctions and density of states (DOS) of SQSs with $n = 2 \sim 7$, respectively. In each panel, the first subplot shows the squared modulus of the wavefunctions of the lowest five states (i.e., the probability density), where the vertical axis represents $|\psi(x)|^2$ and the horizontal axis represents the position. The second subplot displays the density of states, with the vertical axis denoting the DOS and the horizontal axis denoting the position. The potential is set to $U_0 = 5$ for all cases.

where m is the electron mass and U_0 is the potential parameter.[35] Taking a one-dimensional binary SQS as an example, the potential is set to $= 0$ at A sites and to $U_0 = 5$ at B sites, thereby constructing a periodic potential field. By solving the time-independent Schrödinger equation under periodic boundary conditions, the eigenfunctions and the density of states of the system are obtained (Fig. 6).

For SQSs with small unit-cell sizes (Figs. 6(a)–6(c)), the density of states still exhibits a pronounced discrete spectral structure and displays features resembling energy bands and quasi-band gaps. This behavior indicates that, at finite cell sizes, the system retains a strong periodic character, and its eigenfunctions remain spatially extended.

As n increases (Figs. 6(d)–6(f)), corresponding to an increasing SQS supercell size, the density of states gradually evolves into a quasi-continuous distribution, while the eigenfunctions exhibit pronounced spatial localization, primarily confined to finite potential-well regions. This evolutionary trend can be understood as follows: SQSs with small supercell sizes can match only low-order correlation functions of a disordered structure and therefore fail to adequately capture the statistical characteristics of the random potential. As n increases, higher-order components of the correlation functions are progressively satisfied, indicating that the SQS struc-

ture gradually approaches a truly disordered system. Consequently, its electronic structure exhibits characteristic features of disorder, including a quasi-continuous density of states and Anderson localization behavior.[36]

In conventional methods, averaging the results over many randomly generated configurations is usually adopted for disordered systems. However, in the Kronig–Penney model, results for random configurations generally exhibit strong configuration dependence and large statistical fluctuations.[37] The SQS method aims to replace the statistical averaging over many random configurations with a single structure that perfectly matches the correlation functions of the disordered system. Because multiple different SQSs can be constructed for a given n, the physical results obtained from different SQSs should be consistent or differ only slightly. This consistency ensures that a single SQS can provide a representative and reliable result for the disordered system.[25]

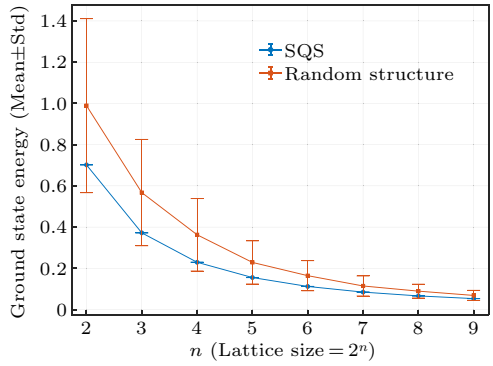

Fig. 7. Average ground-state energy and its standard deviation for ensembles of SQSs and random structures with the same system size 2^n as a function of n. Blue symbols denote SQSs, and orange symbols denote random structures. The horizontal axis represents n, and the vertical axis represents the energy. The potential parameter is $U_0 = 5$. Additionally, due to the extremely small standard deviation of the SQS, its error bars may appear as a single horizontal line.

In the Kronig–Penney model, we compute the mean ground-state energy and its standard deviation for multiple sets of SQSs and for averaging random structures (Fig. 7). For each n, 1000 random structures were sampled for averaging. As the number of available SQS configurations is limited for small n, we specifically used 2, 4, and 32 SQS for $n = 2, 3, 4$, respectively. For $n = 5 \sim 9$, the calculation was performed using 1000 SQS structures to match the sample size used for the random structures. As the cell size increases, the mean ground-state energies of both SQSs and random structures gradually approach zero, indicating that they exhibit consistent lower spectral-edge behavior in the thermodynamic limit. This is consistent with the theoretical expectation for random potential systems: as the system size tends to infinity, the ground-state energy approaches the minimum value of the potential distribution, which is zero in the present model.[38] It is noteworthy that, for the same cell size, the mean ground-state energy of SQS is closer to the limiting value, and the energy fluctuations among different SQS configurations are significantly smaller than those of the random structures. This indicates that, at finite cell sizes, SQS can more efficiently approximate the statistical average of a disordered system and substantially reduce the statistical fluctuations arising from configuration sampling, thus providing a more efficient and stable approximation to disordered structures in finite-size calculations.

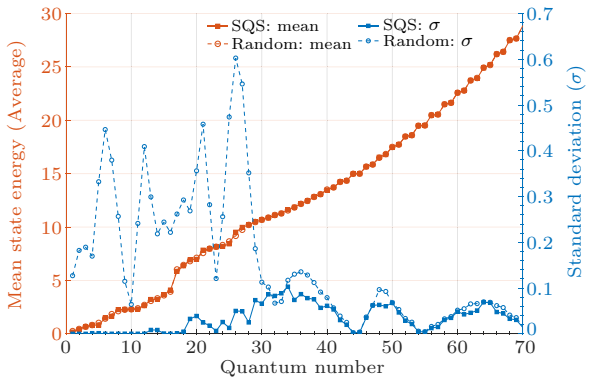

Fig. 8. Mean energy and standard deviation of each energy level for ensembles of SQSs and random structures at $n = 5$. The left vertical axis corresponds to the mean energy of the energy levels (orange symbols), and the right vertical axis corresponds to the standard deviation (blue symbols). Solid lines indicate SQSs, and dashed lines indicate random structures. The potential parameter is $U_0 = 10$.

Further, we calculate the mean energy and the standard deviation of each energy level for different 4096 SQS configurations and random structures (Fig. 8). The results show that, for energies below the barrier height U_0, the statistical fluctuations of the energy levels among different SQS configurations are significantly smaller than those of random structures, whereas for energies above U_0, SQS and random structures exhibit comparable magnitudes of statistical fluctuations. This behavior can be understood from the perspective of wavefunction localization in disordered systems. In the one-dimensional random-potential Kronig–Penney model, eigenstates in the low-energy regime (below the barrier height) are typically characterized by pronounced spatial localization. This behavior can be understood in terms of wavefunction localization in a disordered system. For the one-dimensional Kronig–Penney model with a random potential, the eigenstates in the low-energy regime (below the barrier height) generally exhibit pronounced spatial localization,[39] leading to a strong sensitivity of both their energies and spatial distributions to local atomic configurations. Because SQSs reproduce the statistical features of a disordered system within a finite cell by matching correlation functions of different orders, the statistical fluctuations arising from configurational sampling are significantly suppressed; by contrast, completely random structures still exhibit much larger configuration-to-configuration fluctuations at finite sizes. When the energy exceeds the barrier height U_0, the eigenstates progressively evolve into extended states,[40] whose spatial distributions become much less sensitive to the specific atomic arrangement,

thereby yielding similar statistical behavior for SQSs and random structures. The above results indicate that the SQS method can reliably describe low-energy physical quantities of disordered systems that are sensitive to local configurations, while providing a statistically stable and efficient approximation without averaging over a large number of random configurations.

3.4. Algorithmic efficiency analysis

In general, there are k^m possible configurations for a structure with m sites and k kinds of components. The main idea of generating SQS with high efficiency is the number constraint of many-body clusters. As shown in Fig. 9(a), the simulated annealing algorithm fails to converge to an accurate solution within 3000 s when $n > 8$, whereas the algorithm proposed in this work can obtain solutions for $n \leq 17$ within 300 s. In the algorithm proposed in this work, each step can accurately construct an n-body term (shown in Fig. 9(b)). Consequently, the many-body–constrained algorithm can construct an SQS that perfectly satisfies the prescribed many-body interactions within only a few steps. As shown in Fig. 9(c), during the construction of SQS, the number of n-body terms in the simulated annealing algorithm fluctuates repeatedly. This demonstrates that, by imposing many-body interaction constraints, the algorithm enables extensive pruning of the configuration space, thereby significantly improving computational efficiency.

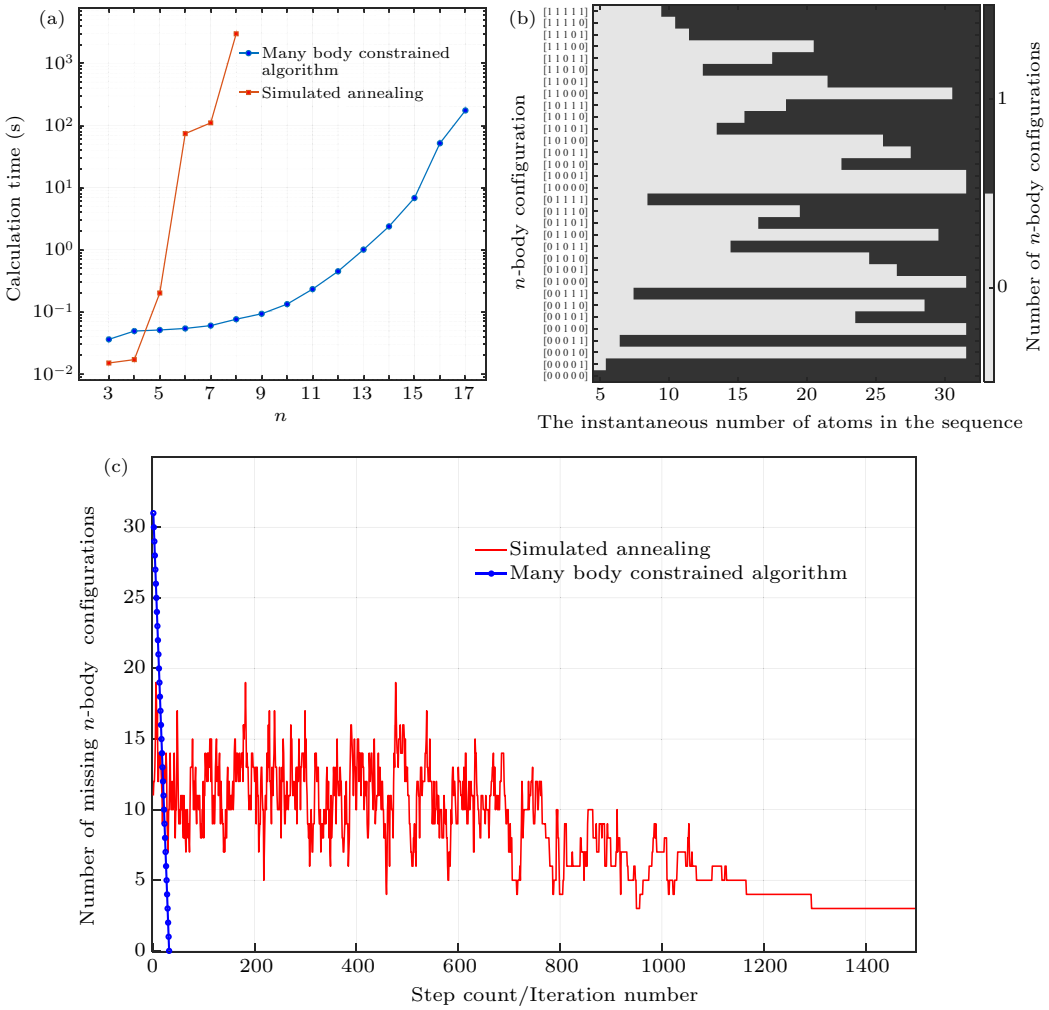

Fig. 9. (a) Comparison of computational efficiency between the proposed many-body-interaction-constrained algorithm and the simulated annealing algorithm for a binary SQS. The computational time required to evaluate $B(2,n)$ is plotted as a function of n. (b) Statistics of n-body cluster occurrences at each construction step of the binary $n = 5$ ($B(2,5)$) SQS generation. The vertical axis represents different n-body cluster configurations, with gray and black cells indicating their absence or single occurrence within the sequence, respectively. (c) Efficiency comparison in constructing a binary $n = 5$ SQS. The number of missing n-body clusters is shown relative to the computational steps of the many-body-interaction-constrained algorithm and the iterations of the simulated annealing algorithm.

4. Conclusion

In summary, we have proposed a method for rapidly constructing special quasirandom structures (SQSs) that perfectly reproduce the prescribed many-body correlations. We have successfully constructed binary SQSs matching up to 22-body interactions, ternary SQSs matching up to 12-body interactions, and SQSs for systems with higher numbers of components. We further demonstrate that the constructed SQSs

possess several advantageous properties: they simultaneously satisfy all $1 \sim n$-body correlations, reproduce two-body correlations up to order $1 \sim (n-1)$ (for binary SQSs), and exhibit a well-defined and universal counting rule for consecutive identical-atom sequences. Using the Kronig–Penney model, we find that the energies of both SQSs and random structures gradually converge to the same thermodynamic limit as the cell size increases. This behavior indicates that the SQS approach does not introduce any systematic bias. Meanwhile, at a given cell size, SQSs show significantly reduced configuration-induced fluctuations in low-energy level statistics, indicating that the SQS approach provides an efficient finite-size approximation to disordered systems. Energy-level statistics and wavefunction analysis indicate that eigenstates are strongly localized below the potential energy, and thus their energies are highly sensitive to the local atomic arrangement. By matching the correlation functions of disordered structures, SQSs suppress finite-size sampling fluctuations in this regime. At high energies, eigenstates become increasingly extended and therefore less dependent on the specific atomic arrangement, resulting in similar statistical behavior for SQSs and random structures. Therefore, the SQS method developed in this work generates high-quality SQSs with superior performance. At finite cell sizes, it provides a stable and reliable description of low-energy electronic properties in disordered systems that are sensitive to local atomic environments, thereby offering a solid physical foundation for efficient electronic-structure calculations of disordered materials.

Program availability

The source code is openly available in Science Data Bank at https://doi.org/10.57760/sciencedb.j00113.00316.

Acknowledgements

This work is financially supported by the National Natural Science Foundation of China (Grant Nos. 12474228 and 12504268) and the Fund of State Key Laboratory of Quantum Functional Materials (Grant No. QFM2025KF008).

References

[1] Johansson C H and Linde J O 1935 *Ann. Phys. (Berlin)* **417** 1
[2] Anderson P W 1958 *Phys. Rev.* **109** 1492
[3] Popescu V and Zunger A 2010 *Phys. Rev. Lett.* **104** 236403
[4] Jankousky, M, Pashov D, Mazo J H, *et al.* 2026 *Nat. Phys.* **22** 88
[5] Sanchez J M, Ducastelle F and Gratias D 1984 *Physica A* **128** 334
[6] Tian F, Lin D Y, Gao X, *et al.* 2020 *The J. Chem. Phys.* **153** 034101
[7] Schoen J M 1969 *Phys. Rev.* **184** 858
[8] Soven P 1967 *Phys. Rev.* **156** 809
[9] Zunger A, Wei S H, Ferreira L G, *et al.* 1990 *Phys. Rev. Lett.* **65** 353
[10] Kohn W and Sham L J 1965 *Phys. Rev.* **140** A1133
[11] Van De Walle A 2009 *Calphad* **33** 266
[12] Jiang C and Uberuaga B P 2016 *Phys. Rev. Lett.* **116** 105501
[13] Shin D, van de Walle A, Wang Y, *et al.* 2007 *Phys. Rev. B* **76** 144204
[14] Sahara R, Emura S, Ii S, *et al.* 2014 *Sci. Technol. Adv. Mater.* **15** 035014
[15] Shin D, Arróyave R, Liu Z K, *et al.* 2006 *Phys. Rev. B* **74** 024204
[16] Wei S H, Ferreira L G, Bernard J E, *et al.* 1990 *Phys. Rev. B* **42** 9622
[17] Yamada R, Masago A, Fukushima T, *et al.* 2021 *Solid State Commun.* **323** 114115
[18] Jiang B, Lin D Y, Wang X, *et al.* 2022 *J. Appl. Phys.* **132** 224101
[19] Jiang Z, Nahas Y, Xu B, *et al.* 2016 *J. Phys.: Condens. Matter* **28** 475901
[20] Voas B K, Usher T M, Liu X, *et al.* 2014 *Phys. Rev. B* **90** 024105
[21] Wang H, Chroneos A, Jiang C, *et al.* 2012 *Phys. Chem. Chem. Phys.* **14** 11737
[22] Jiang C and Chroneos A 2018 *Phys. Chem. Chem. Phys.* **20** 1173
[23] Wong Z M, Tan T L, Yang S W, *et al.* 2018 *J. Phys.: Condens. Matter* **30** 485402
[24] Wang H, Chroneos A, Jiang C, *et al.* 2013 *Phys. Chem. Chem. Phys.* **15** 7599
[25] van de Walle A, Tiwary P, *et al.* 2013 *Calphad* **42** 13
[26] Shinohara K, Seko A, Horiyama T, *et al.* 2021 *Phys. Rev. Mater.* **5** 113803
[27] Jiang C and Uberuaga B P 2016 *Phys. Rev. Lett.* **116** 105501
[28] Zhang J, King D J, Burr P A, *et al.* 2025 *J. Nucl. Mater.* **620** 156308
[29] Sivardière J and Lajzerowicz J 1975 *Phys. Rev. A* **11** 2090
[30] de Bruijn N G 1946 *Proc. K. Ned. Akad. Wet.* **49** 758
[31] Hierholzer C and Wiener C 1873 *Math. Ann.* **6** 30
[32] Shiu W 1997 *Ars Combin.* **47** 33
[33] Hurlbert G and Isaak G 1994 *Contemp. Math.* **178** 153
[34] Hurlbert G and Isaak G 1993 *J. Comb. Theory Ser. A* **64** 50
[35] Kronig R D L and Penney W G 1931 *Proc. R. Soc. A* **130** 499
[36] Mott N F and Twose W D 1961 *Adv. Phys.* **10** 107
[37] Luna-Acosta G A, Izrailev F M, Makarov N M, Kuhl U, *et al.* 2009 *Phys. Rev. B* **80** 115112
[38] Bishop M, Borovyk V and Wehr J 2015 *J. Stat. Phys.* **160** 151
[39] Ishii K 1973 *Prog. Theor. Phys. Suppl.* **53** 77
[40] Kramer B and MacKinnon A 1993 *Rep. Prog. Phys.* **56** 1469