The N^2-Queens problem

Gal Gantar, Miquel López Escoriza, Alba Carballo Castro

I. Problem setup

A. Introduction

The classical N-Queens problem asks whether N queens can be placed on an $N \times N$ chessboard without attacking each other. In this project, we study a higher-dimensional extension: the *3D N^2-Queens problem*. The board is an $N \times N \times N$ grid, and the goal is to place N^2 queens with the minimum possible number of conflicts. Since the configuration space of size $\binom{N^3}{N^2}$ grows extremely quickly, exhaustive search is infeasible for all but small sizes. We therefore rely on Markov Chain methods to explore the space and search for low-conflict configurations[1].

B. State space

We place $Q = N^2$ queens on an $N \times N \times N$ board. A state is defined as the set of occupied cells,

$$S = \{(i_q, j_q, k_q) \in \{0, \dots, N-1\}^3 : q = 1, \dots, Q\},$$

with the constraint $|S| = Q$ so that no two queens share a position.

C. Objective function

To evaluate a configuration, we define *attacking* between two queens. Queens in positions (i_1, j_1, k_1) and (i_2, j_2, k_2), are attacking each other if:

1) they share two coordinates (i.e., lie on an axis-aligned line);
2) they lie on a 2D diagonal in any coordinate plane;
3) they lie on a 3D space diagonal, i.e. $|i_1 - i_2| = |j_1 - j_2| = |k_1 - k_2|$.

Let $E(S)$ be the energy of state S denoting the number of attacking pairs of queens in the state. The optimization goal is then $\min_S E(S)$.

II. Theoretical background

Previous work has mainly focused on determining $Q_{\max}(N, 3)$, the largest number of non-attacking queens that can be placed on an $N \times N \times N$ cube. We are interested in existing constructions from the literature and examine how they can be adapted to build strong partial or complete solutions to our problem.

A. Construction of solutions

a) Maximum number of queens: Two queens may not share two coordinates. Since there are only N^2 distinct ordered pairs (i, j), any valid placement satisfies

$$Q_{\max}(N, 3) \leq N^2.$$

Thus, the central question is whether this upper bound can be achieved for a given N.

[1]Code publicly available on https://github.com/galgantar/monte-carlo-collective

b) Layer constraint: In any optimal configuration (if it exists), no two queens may share the same pair of first two coordinates (i, j). Thus, each (i, j) must be occupied by exactly one queen with an assigned height $k_{i,j}$.

c) Known results: Table 1 lists the values of $Q_{\max}(N, 3)$ obtained computationally in [1]. None of these reach the theoretical upper bound discussed above, highlighting the relevance of constructing strong partial solutions.

N	3	4	5	6	7	8	9	10
$Q_{\max}(N, 3)$	4	7	13	21	32	48	67	91

Table 1: Best known values of the maximum number of non-attacking queens in 3D, for $N = 3$ to 10, following [1].

For the specific values $N \leq 10$, we include arguments for non-existance of the valid configuration in the Appendix A.

B. Valid solutions

Theorem II.1 (Klarner [2])**.** *If $\gcd(N, 210) = 1$, there exists a configuration of N^2 non–attacking queens on the N^3 cube.*

A valid construction is obtained by placing queens in all positions (i, j, k) satisfying

$$k \equiv 3i + 5j \pmod{N}.$$

If $\gcd(N, 210) \neq 1$ the existence of valid solutions for large N remains an open question.

C. Partial solutions

When a full N^2 configuration is not known or does not exist, we try to construct placements with few conflicts. Note that there is no theoretical justification that the following solutions yield the optimal partial solution.

a) Latin-square partial solution: A Latin square of order N is an $N \times N$ array whose entries are taken from $\{0, \dots, N-1\}$ and where each value appears exactly once in every row and column [3]. A latin square of order N can be defined by

$$L_{ij} = (i + j) \bmod N.$$

Placing N^2 queens on coordinates (i, j, L_{ij}) produces placements that satisfy all axis-alignment constraints. Diagonal conflicts may still occur, but the structure of the Latin square ensures that queens are spread regularly across layers.

b) Klarner-style partial solution: If $\gcd(N, 210) = 1$, Klarner's construction II.1 provides a valid solution. For values of N that do not satisfy this constraint, we use the same rule on the largest subcube of dimension N' with $\gcd(N', 210) = 1$ and fill the remaining positions randomly. This produces a configuration where a large subset of queens is already conflict-free, resulting in a low number of conflicts.

III. Proposed approach

A. Metropolis–Hastings algorithm

We develop an MCMC procedure for sampling configurations of N^2 queens on an $N \times N \times N$ board, under the *layer constraint* introduced in Paragraph II-A0b, which enforces that each vertical file (i, j) contains exactly one queen.

This restriction reduces the configuration space while preserving all known optimal constructions. However, it is not proven that restricting the dynamics to this subset of configurations yields an optimal solution for all values of N.

Under the layer constraint, each configuration is fully described by a height array

$$S = (k_{i,j})_{1 \le i,j \le N}, \qquad k_{i,j} \in \{1, \dots, N\},$$

so that the queen indexed by (i, j) occupies the cell $(i, j, k_{i,j})$.

a) Base chain ψ: The base chain modifies the height of one queen at a time. A state T is a neighboring state of S if it differs from S in the position of exactly one queen: if $\exists!(i,j) \quad k_{i,j}(T) \neq k_{i,j}(S)$, we write $T \sim S$. Given a state S, the chain performs:

1) select (i, j) uniformly from the N^2 possible pairs,
2) choose a new height $k' \in \{1, \dots, N\} \setminus \{k_{i,j}\}$ uniformly,
3) form T by setting $k_{i,j}(T) = k'$ and leaving all other heights unchanged.

Each configuration has exactly $N^2(N-1)$ neighbors, and every neighbor is proposed with probability

$$\psi_{ST} = \begin{cases} \dfrac{1}{N^2(N-1)} & T \sim S, \\ 0 & \text{otherwise.} \end{cases}$$

The proposal chain is irreducible, symmetric ($\psi_{ST} = \psi_{TS}$) and aperiodic ($p^{(2)}_{SS} > 0$ and $p^{(3)}_{SS} > 0$ for each S).

b) Calculation of energy function: The energy $E(S)$ is the number of attacking pairs of queens.

For configuration $S = (k_{i,j})_{1 \le i,j \le N}$, we define:

$$I = (i)^N_{i,j=1}, \qquad J = (j)^N_{i,j=1}, \qquad K = (k_{i,j})^N_{i,j=1}.$$

Using the vectorizations of matrices I, J, K and the relations described in I-C we can efficiently compute:

$$C_S(i,j) = \#\{\text{conflicts queen on } (i,j) \text{ has in state } S\}.$$

By the definition of the base chain, a move from S to T changes a single height of the queen at (i, j) from $k_{i,j} \to k'$. All other queens remain fixed, so only the conflicts involving this queen change, and the energy difference can then be computed locally as:

$$\Delta E = E(T) - E(S) = C_T(i,j) - C_S(i,j).$$

The vectorized calculation of the energy can be computed in $\mathcal{O}(N^2)$ per move.

c) Target distributions: We seek to sample preferentially from low-energy configurations. The ideal target distribution π_∞ places a uniform mass on the set of configurations, reaching the minimum achievable energy,

$$\pi_\infty(S) = \begin{cases} \dfrac{1}{Z_\infty} & \text{if } E(S) = E^*, \\ 0 & \text{otherwise,} \end{cases} \qquad E^* = \min_{T \in \mathcal{S}} E(T).$$

As direct sampling is infeasible, we use the "inverse temperature" β to define the distributions π_β:

$$\pi_\beta(S) = \frac{e^{-\beta E(S)}}{Z_\beta}, \qquad Z_\beta = \sum_{T \in \mathcal{S}} e^{-\beta E(T)}.$$

The uniform distribution arises at $\beta = 0$, and $\pi_\beta \to \pi_\infty$ as $\beta \to \infty$.

d) Metropolis–Hastings acceptance probabilities: Given the proposal chain ψ and a target distribution π_β, the Metropolis–Hastings transition probabilities are

$$P_{ST} = \begin{cases} \psi_{ST}\, a_{ST} & S \neq T, \\ 1 - \displaystyle\sum_{U \neq S} \psi_{SU} a_{SU} & S = T, \end{cases}$$

where

$$a_{ST} = \min\left(1, \frac{\pi_\beta(T)\, \psi_{TS}}{\pi_\beta(S)\, \psi_{ST}}\right)$$

are the acceptance probabilities. Since ψ is symmetric, this simplifies to

$$a_{ST} = \min\left(1, e^{-\beta(E(T) - E(S))}\right).$$

If $E(T) \le E(S)$ the move is always accepted; otherwise it is accepted with probability $e^{-\beta(E(T)-E(S))}$. The resulting chain is irreducible, aperiodic, positive recurrent, and therefore ergodic with stationary distribution π_β.

B. Simulated annealing

Simulated annealing extends the basic Metropolis–Hastings approach by allowing the inverse temperature β to vary over time. Rather than sampling from a fixed distribution π_β, the chain is run with a time-dependent inverse temperature $\beta(t)$.

At low values of $\beta(t)$, the chain behaves similarly to a random walk and can freely move across state space, which helps escape local minima. As $\beta(t)$ increases, the stationary distribution becomes increasingly concentrated in low-energy configurations, biasing the dynamics toward solutions with fewer conflicts.

The choice of how $\beta(t)$ evolves is known as the *cooling schedule*. Common schedules include linear, exponential, and logarithmic annealing, which trade off exploration against convergence stability.

C. Initialization

We initialize the Markov chain from the structured partial solutions introduced in Section II-C, together with a purely random baseline:

- Latin-square initialization: start from the Latin-square construction (Section II-C), which assigns one queen to

each (i, j) with height L_{ij} and already satisfies the axis constraints.
- Klarner initialization: when $\gcd(N, 210) = 1$, use Klarner's construction as an exact starting solution; otherwise embed the largest admissible subcube and fill the remaining positions at random, as described in Section II-C.
- Random initialization: place Q queens uniformly at random heights for each pair of coordinates (i, j), providing an unstructured baseline.

These three schemes allow us to compare how structured partial solutions affect convergence speed and the quality of the final energy.

IV. Results and Discussion

In the following, we present the different results from our experiments.

A. Finding the best range for β

First, we explore how energy changes as a function of time. As an initial approach, we choose $N = 12$ and run the simulation 20 times for 1.000.000 steps with random initialization of the board and linear annealing. We tested different ranges $(\beta_{start}, \beta_{end}) \in \{(0.5, 3), (0.1, 5), (1, 3), (1, 5)\}$. Figure 1 shows the energy history and its standard deviation for the different ranges. Figure 2 shows the acceptance rate history.

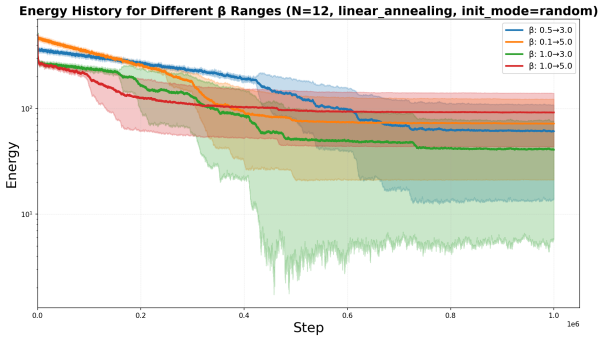

Fig. 1: Energy history for fixed N and different ranges of β.

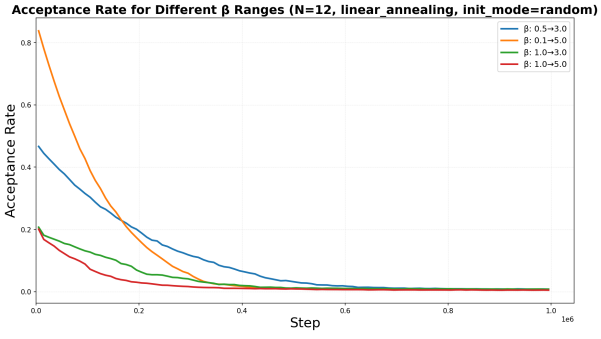

Fig. 2: Observed acceptance probabilities for fixed N and different ranges of β.

From both plots, we observe that the range $(\beta_{\text{start}}, \beta_{\text{end}}) = (1, 3)$ provides the most balanced behavior. In this case, the chain is able to explore the state space during the early iterations while still converging toward lower-energy configurations later in the run.

For the wider ranges $(0.1, 5)$ and $(0.5, 3)$, the exploratory phase is excessively long: acceptance probabilities remain high for most of the simulation, preventing the chain from sufficiently concentrating on low-energy states. By contrast, for $(1, 5)$ the exploratory phase is too short, and many moves are rejected early on. As a result, the Markov chain quickly becomes trapped in a local minima.

B. Finding the best schedule

Once the optimal range has been determined, we decided to further explore the effect of simmulated annealing. For that, we test keeping $\beta = 3$ as a constant or using simmulated annealing with linear, exponential, sinusoidal and logarithmic schedules, as depicted in Figure 3. We choose $N = 12$ and perform 20 different runs for a million steps.

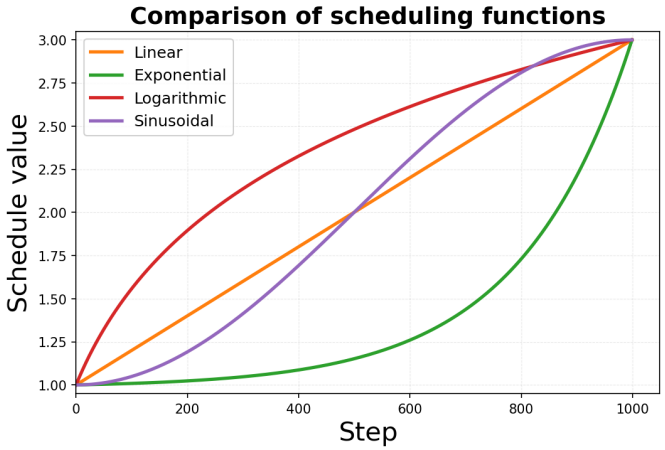

Fig. 3: Comparison between scheduling functions in simmulated annealing.

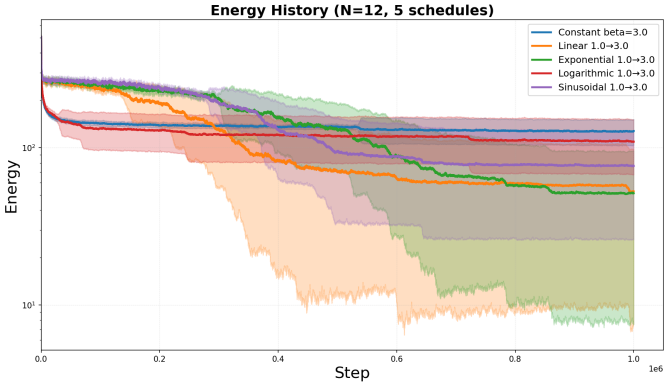

Fig. 4: Energy history across different schedules for $N = 12$ and β range $[1, 3]$.

The results in Figure 4 show that although constant and logarithmic schedules initially appear promising, they eventually get trapped in local minima, while linear, exponential and sinusoidal schedules are able to escape them, with the linear schedule reaching lowest energies on average. This can be

explained by the evolution of β shown in Figure 3: logarithmic and sinusoidal schedules increase β too quickly in the early phase, rapidly suppressing uphill moves and causing premature trapping. In contrast, linear and exponential schedules keep β low for longer, allowing broader exploration before gradually enforcing convergence, which leads to better final energies.

C. Energy as a function of N

Furthermore, we wanted to explore how the minimal energy reached changed as a function of the board size N. Using different initialization techniques, we report the minimal energy reached as the mean and standard deviation over 20 different runs for linear annealing, 5 million steps, and values of $N \in \{2, 3, \ldots, 15\}$.

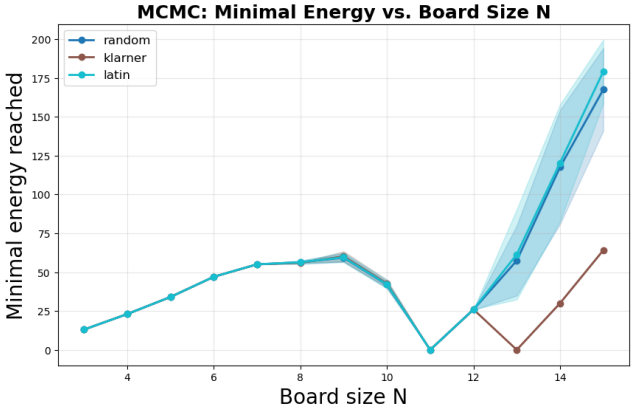

Fig. 5: Minimal energy for different value of N.

In particular, Figure 5 shows that the minimal energy reached grows with N in a similar way for the three initializations until $N = 8$ where it starts to decline until $N = 11$ where all three initializations reach zero energy. From $N = 13$, we see a clear advantage in using the Klarner initialization, reaching energy 0 for $N = 13$ but also noticeably improving the minimal energy for $N = 14, 15$ regardless of the fact that $14, 15$ are not coprime with 210.

In addition, we also computed the mean and standard deviation of the number of steps needed to reach the optimal energy for each of the runs, which is reported in Figure 6. In particular, we see that the number of steps increases until $N = 6$ and then begins to descend again until $N = 11$. For bigger values, the number of necessary steps goes above 3 and 4 million on average. Again we can see the advantage of using Klarner initialization even for the values of N not coprime with 210.

V. Increasing N

Our method does not explicitly adapt its parameters to the size of the board N. However, as N increases, both the energy landscape and the size of the state space increase, affecting the convergence. As shown in Figure 6, larger values of N generally require more iterations for the algorithm to reach near-optimal configurations.

This effect is also reflected in the choice of the inverse-temperature range. In Figure 7, we compare two ranges, $[1, 3]$ and $[1, 5]$, for board sizes $N = 12$ and $N = 15$. For $N = 12$, the smaller range performs better, whereas for $N = 15$ the larger range leads to improved results. This suggests that the final inverse temperature β_{end} should increase with N to maintain good performance as the size of the problem increases.

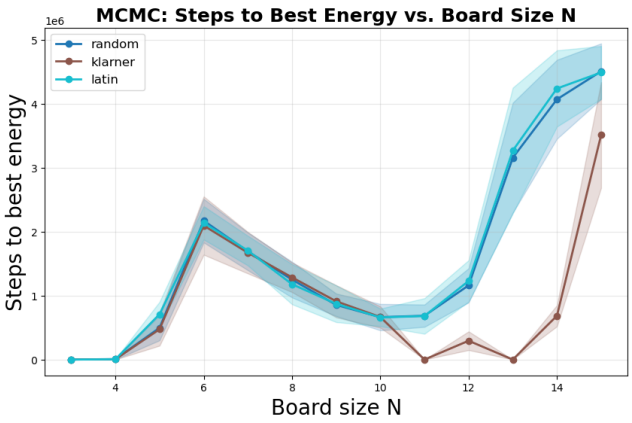

Fig. 6: Steps to best energy for different value of N

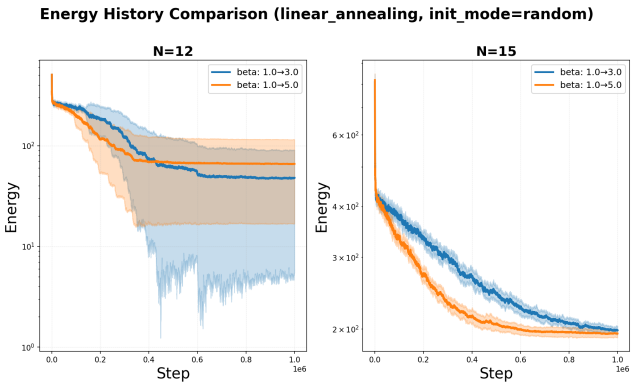

Fig. 7: Steps to best energy for different value of N.

A. How should β scale with N?

In this section, we provide theoretical estimates on how the inverse temperature β and the number of MCMC steps should scale with N.

Let $f(x)$ denote the energy of state x. Consider the distribution π_β as defined above. Let $\mathcal{M}_0$ denote the set of configurations with minimum energy f_0 and $\mathcal{M}_1$ set of configurations with second minimum energy $f_1 = \min_{x, f(x) \neq f_0} f(x)$. A standard approximation with $N_0 = |\mathcal{M}_0|$ and $N_1 = |\mathcal{M}_1|$ gives:

$$\pi_\beta(\mathcal{M}_0) \approx \frac{1}{1 + \frac{N_1}{N_0} e^{-\beta(f_1 - f_0)}}.$$

Requiring $\pi_\beta(\mathcal{M}_0) = 1 - \varepsilon$ leads to

$$\beta \approx \frac{1}{f_1 - f_0} \log\left(\frac{N_1}{\varepsilon N_0}\right). \tag{1}$$

Although N_0 and N_1 are unknown in the 3D N–Queens problem, it is reasonable to assume that the number of

minimal-energy configurations grows polynomially relative to the number of second-minimal-energy ones,

$$\frac{N_1}{N_0} \approx N^d, \qquad d > 0.$$

Substituting this into (1) yields:

$$\beta \approx \frac{d \log N}{f_1 - f_0} + \mathcal{O}(1),$$

so the inverse temperature required increases only logarithmically in N. If we further assume that $f_1 - f_0 = \mathcal{O}(1)$[2] we get that the value of β should scale:

$$\beta_{\text{end}}(N) \propto \log N.$$

However, for the exploratory phase, no such scaling is needed. To keep the acceptance probability for uphill moves,

$$\exp(-\beta\, \Delta f),$$

roughly constant across different sizes of the board, the product $\beta\, \Delta f$ should remain stable. In random configurations, a queen attacks $O(N)$ squares but the queen density is $1/N$, so the expected number of conflicts involving a queen is $O(1)$; therefore, moving a queen changes the number of conflicts by an amount whose distribution does not depend on N. Thus,

$$\mathbb{E}[\Delta f \mid \Delta f > 0] = \Theta(1).$$

Maintaining a constant acceptance probability then requires $\beta = \Theta(1)$ during the exploratory phase. Consequently, the algorithm can be initialized with an inverse temperature independent of N,

$$\beta_{\text{start}} = \Theta(1).$$

B. How should the number of MCMC steps scale with N?

Starting from the mixing-time inequality for an ergodic Markov chain,

$$T_\varepsilon \;\le\; \frac{1}{\gamma_N} \left(\log \frac{1}{\varepsilon} + \frac{1}{2} \log \frac{1}{\pi_{\min}} \right),$$

we estimate how each term scales with N. The state space has size

$$|\Omega_N| = N^{N^2}, \qquad \log |\Omega_N| = \Theta(N^2 \log N),$$

and the stationary distribution is roughly spread over this set, so we take

$$\log \frac{1}{\pi_{\min}} \approx \log |\Omega_N| \approx N^2 \log N.$$

Substituting into the bound gives

$$T_\varepsilon \lesssim \frac{1}{\gamma_N} N^2 \log N.$$

If we assume that the spectral gap decays polynomially, $\gamma_N \gtrsim N^{-\alpha}$, then

$$T_\varepsilon = O(N^{\alpha+2} \log N).$$

[2]There is no known proof that validates this assumption. It is also possible that the difference depends on N (and it can be shown that it cannot grow faster than $\Theta(N)$). If this was the case, we would have to adjust our estimation for the (unknown) factor of growth.

VI. Conclusion

In this project, we studied the 3D N^2-Queens problem through the lens of Markov Chain Monte Carlo methods. By imposing one queen per pair of the first two coordinates (i, j), we reduced the configuration space and designed a Metropolis–Hastings sampler combined with simulated annealing. This made it possible to explore board sizes beyond what brute force could handle and to compare different annealing schedules and initialization strategies.

We found that the choice of schedule matters: overly aggressive temperature ranges tend to freeze the chain too early, whereas a moderate linear annealing schedule was the most consistent at escaping local minima and reaching low energies. Initialization also had an impact. Klarner-based initialization generally converged faster and achieved better final energies than random or Latin-square baselines, even in cases where Klarner-based initialization does not provide an optimal solution.

At the same time, our results highlight the limits of this approach. For many values of N, the chain still gets trapped in local minima, and we do not have guarantees that the method reaches the global optimum. The scaling discussion provides useful intuition for choosing β ranges and run lengths, but it remains a heuristic.

Overall, this work shows that MCMC with simulated annealing, paired with problem-specific initializations, is a practical way to obtain low-conflict configurations in the 3D N^2-Queens problem. Future improvements could focus on better move proposals, faster incremental energy updates, and more adaptive annealing strategies to improve robustness as N grows.

References

[1] T. Kunt, "The n-queens problem in higher dimensions," 2024. 1, 6

[2] D. A. Klarner, "The problem of embedding knight's tours and queens' solutions in higher dimensions," *Discrete Mathematics*, vol. 22, no. 2, pp. 147–151, 1978. 1

[3] B. D. McKay and I. M. Wanless, "On the number of latin squares," 2009. 1

[4] V. Chvátal, "Colouring queen graphs." http://www.cs.mcgill.ca/~cs/GI/chvatal/. Accessed 2024. 6

APPENDIX A
NON-EXISTENCE OF PERFECT SOLUTIONS FOR $N \leq 10$

In this appendix, we summarize why no valid configuration of N^2 queens on $N \times N \times N$ board exists for $N \leq 10$.

A. The cases $N = 2, 4, 6$

A $2 \times 2 \times 2$ subcube can contain at most one non-attacking queen, since any queen placed inside attacks all other seven cells via axis, planar, or space diagonals. When N is even, the N^3 cube partitions into

$$(N/2)^3$$

disjoint 2^3 subcubes, giving the upper bound

$$Q \leq \frac{N^3}{8}.$$

For $N = 2, 4, 6$ this upper bound is strictly smaller than N^2:

$$N = 2 : \ 1 < 4, \qquad N = 4 : \ 8 < 16, \qquad N = 6 : \ 27 < 36,$$

so a perfect configuration with N^2 queens cannot exist.

B. The cases $N = 8, 9, 10$

Following Chvátal's queen-graph formulation [4], the classical 2D queen problem corresponds to finding an n-colouring of the queen graph $G_{n,2}$. Kunt [1], based on results of Chvátal, shows that no such n-colouring exists for $n = 8, 9, 10$.

Since an n-colouring of $G_{n,2}$ is a necessary condition for the existence of a valid configuration of N^2 queens on N^3 boeard (Corollary 4 and Corollary 7 of [1]), we can conclude that none of $N = 8, 9, 10$ admits a full solution.

C. The cases $N = 3, 5, 7$

Finally, for $N = 3, 5, 7$, Kunt's computational results [1] show that the maximal value $Q_{\max}(N, 3)$ is strictly smaller than N^2. Therefore, no perfect configuration exists for these values either.