OXFORD JOURNAL
OF STUDENT
SCHOLARSHIP

Exact Mixing Time of the Corner-Permutation Random Walk on the 2×2×2 Rubik's Cube

Jaiveer Gulyani
gulianijai@gmail.com

ABSTRACT

How many quarter turns are needed on average to scramble a Rubik's cube? This paper gives an exact answer for the corner-permutation projection of the 2×2×2 cube as a random walk on 40,320 permutations of eight corners. The walk of exactly k moves cannot get closer to a uniform distribution on the 40,320 states than the walk of (k−1) moves. We thus measure the total variation (TV) distance to the uniform distribution on the correct parity class. After 10 quarter turns the TV distance is down to 0.231 or 23.11 % from 0.5; after 20 quarter turns it is down to 0.187 or 18.7 % from 0.5 (computed by iterating the full probability vector instead of sampling it). The TV-distance curve starts with a plateau followed by geometric decay at the rate given by the spectral gap; hence the transition window is of about the same length as the mixing time itself. Removing one face pair from the generating set increases the mixing time by 55% as the spectral gap predicts. We applied the same method to the full corner group, orientation inclusive, of 88,179,840 states. The mixing time for the 2×2×2 cube under the 18-move chain is 19 while for the permutation projection it is 11. Corner permutation thus accounts for about 11 of the 19 moves; orientation for the remaining 8 moves.

Keywords: Markov chains; mixing time; Rubik's Cube; random walks on groups; total variation distance; spectral gap; lazy random walk; cutoff phenomenon

1. INTRODUCTION

The Rubik's Cube is a toy, but it is also a finite group: each configuration is a group element and each face turn is a generator (Singmaster, 1981; Chen, 2004; Joyner, 2008). This turns the everyday question "how scrambled is this cube?" into a precise question about random walks in groups.If we start with a cube that is solid and randomly and uniformly pick to toss it (independently) then how many tosses of the cube will it take for the resulting configuration of faces to be (statistically) essentially uniformly distributed among all possible configurations of faces?

The standard way to measure this is the total variation (TV) distance between the walk's distribution after t steps and the uniform distribution. The mixing time is the first step at which this distance falls below a fixed threshold, conventionally 1/4 (Levin et al., 2017). For some families of random walks the TV distance stays close to 1 for a long time and then drops abruptly. This is the cutoff phenomenon, first

observed for random transpositions by Diaconis and Shahshahani (1981), named by Aldous and Diaconis (1986) following Aldous's (1983) study of random walks on finite groups, and surveyed by Diaconis (1996) and Saloff-Coste (2004). Cutoff is an asymptotic property of a *sequence* of chains whose state spaces grow (Levin et al., 2017, Ch. 18), so a single finite chain cannot be said to have or lack it.For a single chain, there is a shape, viz. the length of the initial part which is flat, followed by the rate of decay of that flat part. For the kind of exact analysis of a shape such as this paper does, there is a classical model. Bayer and Diaconis in their 1992 paper calculated that seven riffle shuffles are enough to mix a 52-card deck. The results were obtained from the exact distribution of the shuffles, not from samples. The representation-theoretic methods used in this paper are developed in Diaconis's 1988 paper.

For the standard 3×3×3 cube, which has about 4.3×10^{19} configurations, the mixing time is not known. Qu et al. (2024) proved that at least 26 random moves are needed under the 18-move chain, in which a uniformly chosen face is turned by a quarter, half or three-quarter turn. Their method maps each configuration to its distance from solved (a statistic with 21 values). The TV distance can't increase under a map, so the time it takes for the distance statistic to mix must be less than or equal to the time it takes for the full statistic to mix. We estimated the distribution for the distance statistic by sampling one million cubical regions (all solved on a supercomputer to be optimal) per time step, and arrived at a lower bound, which is not exact and not even a group quotient. Fernique (2025) in his follow-up work also determines the 2×2×2 cube (or eight corner pieces of the 3×3×3) to have a mixing time of 19 steps under the 18-step long chain of 18 moves. The 3×3×3 diameter was lowered from 52 moves in Thistlethwaite's 1981 algorithm (Singmaster, 1981) to 26 (Kunkle & Cooperman, 2007), 22 (Rokicki, 2010) and finally exactly 20 in the half-turn metric (Rokicki et al., 2014). The quarter-turn-metric diameter is 26 (Rokicki & Davidson, 2014; see also Rokicki, 2014). Hirata (2024) gives probabilistic estimates of these diameters, not proofs.

This paper studies a system small enough to be computed exactly: the permutation of the eight corners of the 2×2×2 cube, ignoring how each corner is twisted. It has 8! = 40,320 states, so the complete probability vector can be updated at every step. It is important to be precise about how this system relates to the full cube. It is not a subset of the full cube's configurations but a *projection* of them: every configuration of the full cube, held in a fixed spatial frame, has exactly one corner permutation. Because TV distance cannot increase under a projection, the mixing time of the projection is a lower bound on the mixing time of the full cube (Section 2.5).

Research question: How many random quarter turns of a corner do you need to make on a 2×2×2 cube in order to get its corner permutation within TV distance 1/4 of a uniform distribution? How does the answer to this question change for the different ways of defining "number of moves" of a walk on a corner of the cube, and what fraction of the mixing time of the full cube does this permutation account for?

Hypothesis: Random walk on a finite group mixing at a geometric rate given by the second largest eigenvalue of the transition matrix for the group. Thus for a short time the TV curve would be expected to remain roughly constant as the walk explores the group prior to being able to visit a large proportion of

states. Subsequently the curve would be expected to fall roughly at a constant rate determined by the spectral gap.

Contributions: (i) The TV curve is exactly computed for the corner-permutation walk for the three different "number of moves" definitions: for an exact k-move walk, for a lazy chain, and for a continuous-time chain. For all three definitions the mixing time of the corner-permutation walk is about 10 quarter turns. (ii) The spectral behavior of the curve, and the importance of the generating set for the walk, are explained in detail. (iii) The full, orientation-inclusive corner group of size 88,179,840 states is exactly computed by the same method of computation as before. This computation reproduces the previously computed value of 18 moves for the 18-move chain, and also separates the value for the 18-move chain into the value for the permutation part (11 moves) and the value for the orientation part (8 moves). (iv) Parity and "number of moves" are explained in more detail.

Section 2 describes the model and methods, Section 3 reports the results, Section 4 discusses them and their limitations, and Section 5 concludes.

2. METHODS

2.1 State, labelling and generators

The eight corner positions are 0 = URF, 1 = UFL, 2 = ULB, 3 = UBR, 4 = DFR, 5 = DLF, 6 = DBL and 7 = DRB. A position on the cube is denoted by an 8-tuple σ, which in turn is an 8-tuple of values, the value at position j is the corner piece at position j. The solving problem is defined as the identity (0, 1, …, 7). In this description the cube is considered in a fixed spatial frame, that is all six faces can be turned and the whole cube is not turned in space. This makes a big difference on 2×2×2 since U and D′ are clearly distinct in this frame while they are simply the whole-cube rotation in the opposite direction on this cake.

The generating set for these moves is the twelve quarter turns about the centers of the faces of the cube, all four ways. As these moves affect positions on the cube, we write that a move m is a permutation φm of the positions of the cube, and for a state σ of the cube we can thus write σ ◦ φm to get the state of the cube after we have applied the move m to σ. Thus the six faces turn in the same way as in the above description of the six generators for these moves: URF→UBR→ULB→UFL (for U face), DFR→DLF→DBL→DRB (for D face), URF→UFL→DLF→DFR (for F face), ULB→DBL→DRB→UBR (for B face), UFL→ULB→DBL→DLF (for L face) and URF→DFR→DRB→UBR (for R face). The other six generators for these moves are the six inverse moves for these six generators. Because both directions of each face are included with equal frequency in the above descriptions of the generators and their inverses, whether or not a particular direction about a face is labeled "clockwise" makes no difference to the results that are reported here.

Every quarter turn is a 4-cycle and therefore an odd permutation. A breadth-first search (BFS) from the solved state reaches all 40,320 = 8! permutations, so the generators produce the full symmetric group S_8 acting on the positions. If a move table had been wrong, this count would almost certainly have come out

different. The same search records each state's distance from solved: 1, 12, 114, 876, 4,931, 12,972, 15,066, 6,300 and 48 states lie at distances 0 to 8. The diameter of this projection is therefore 8. For comparison, the full 2×2×2 cube (3,674,160 positions up to whole-cube rotation) has diameter 14 in the quarter-turn metric and 11 in the half-turn metric. These values have been known from exhaustive search since the early 1980s and were re-verified here by BFS.

2.2 Three ways to count moves

Exact k-move walk. An exact k-move walk is a walk of length k where each move is chosen uniformly at random from the set of four quarter turns. This process results in a permutation with the same parity as k. Hence the support of Pk is a subset of S8 of size at most ½|S8| and Pk is at least TV distance ½ from a uniform distribution on S8 for all k. The only "research question" that this model does not answer "never" is what is the distribution of πk for all k, where πk is the uniform distribution on all S8 permutations of a given parity k. This is the primary distribution studied in this paper.

Lazy chain. At each discrete step, called a *tick*, the walk stays where it is with probability p_{stay} and otherwise applies one of the twelve generators chosen uniformly. The one-tick update is

$$P_{t+1}(s) = p_{stay}P_t(s) + \frac{1-p_{stay}}{12}\sum_{m=1}^{12} P_t\left(\varphi_m^{-1}(s)\right) \tag{1}$$

where the sum runs over the twelve generators. Because the generating set is closed under inverses, the sum over φ_m^{-1} equals the sum over φ_m, and the chain is symmetric with the uniform distribution π as its stationary distribution. Laziness removes the period-2 parity oscillation. After t ticks, however, the number of moves actually made is random, with a Binomial(t, $1 - p_{\text{stay}}$) distribution, so $t(1 - p_{\text{stay}})$ is an expected move count. The default is $p_{\text{stay}} = 0.5$.

Continuous-time chain. Moves occur at the times of a rate-1 Poisson process, so after time s the number of moves made is Poisson(s). This also removes periodicity, and s is directly an expected number of moves.

2.3 Exact iteration and mixing time

The transition table T[i][m] is precomputed once. The algorithm starts from a point mass placed on the solved state. Then in each step, the full 40,320-entry probability vector is updated exactly without any use of random sampling. The TV distance between P and π is

$$TV\left(P_t, \pi\right) = \frac{1}{2}\sum_s \left|P_t(s) - \pi(s)\right| \tag{2}$$

and the mixing time at threshold ε is

$$t_{mix}(\varepsilon) = \left\{t \geq 0 : TV\left(P_{t'}, \pi_t\right) < \varepsilon\right\} \tag{3}$$

Here πt denotes the parity-class target πk for the exact k-move walk and the uniform distribution π for the lazy and continuous-time chains. Note that according to the definition by Levin et al. (2017) the mixing time is defined using ≤ instead of <. All values for the mixing time have been checked for all reported thresholds and are identical. The values for the mixing time are given for the thresholds ε = 0.5, 0.25, 0.1 and 0.01. The value for ε = 0.25 is marked in bold. All values are exact up to double-precision round-off (about 10^{-15}). The total sum of probabilities was checked to stay within 6×10^{-15} of 1 over 200 ticks. This is a simple check for the code and not a rigorous result.

2.4 Spectral quantities

Let K be the non-lazy 12-move transition matrix. Since K is symmetric all its eigenvalues are real and the second largest eigenvalue λ2 has been computed using a sparse eigensolver. We also note that K has an eigenvalue −1, corresponding to the change of parity for every move. The transition matrix for the lazy chain is p_{stay}I + (1 − p_{stay})K. Hence the second largest eigenvalue λ∗ for the lazy chain and its relaxation time t_{rel} = 1/(1 − λ∗are given by p_{stay}+ (1 − p_{stay})λ_2 and Levin et al. (2017, Theorems 12.4 and 12.5) derive bounds for the mixing time in terms of the relaxation time.

$$\left(t_{rel} - 1\right) \ln ln \frac{1}{2\varepsilon} \leq t_{mix}(\varepsilon) \leq t_{rel} \ln ln \frac{1}{\varepsilon \pi_{min}}, \quad t_{rel} = \frac{1}{1-\lambda_*} \tag{4}$$

where π_{min} = 1/40,320 is the smallest stationary probability. The same quantities were computed for every generating set studied.

2.5 Relationship to the full cube

For the cube held in a fixed frame, the full corner group, i.e. the group of corner permutations with possible corner reversals, has 8!·37 = 88,179,840 elements. Each corner permutation corresponds to exactly 37 = 2,187 of these states. So the walk on permutations is just the image of the walk on the full group under the map that forgets corner orientation. Clearly for any map f and distributions μ and ν on a group, TV(μ ◦ f−1, ν ◦ f−1) ≤ TV(μ, ν), since every event about f(X) is an event about X. So the TV distance of the permutation walk cannot exceed that of the full walk. Hence the mixing time of the permutation walk gives a lower bound for the mixing time of the full walk. This same inequality is used by Qu et al. (2024) for their distance statistic. Qu et al. (2024) use the same inequality for their distance statistic. The frequently quoted count of 3,674,160 = 7! · 3^6 counts positions up to whole-cube rotation. Corner permutation is not well defined up to rotation, so the projection argument applies to the fixed-frame group, not to that count.

We could also measure the difference between our approach and the full corner-to-corner corner-to-corner transition on the full cube for the same number of moves. We stored the probability vector as a 40,320 x 2,187 array, where the first index ran over all permutations of the corners, and the second index ran over all possible orientations of the corners. The two values for a given permutation would be processed one after the other by a move, because a move affects the two coordinates of a corner in a completely independent way, given the corner's position. We checked the 18 move chains of Qu et al. (2024) and Fernique (2025), using all possible faces, and all three possible half-turns, as well as the three quarter-turns about a face's center, and the 12 move quarter-turn chain using both the exact k-move (i.e. k-moves of parity class p) definition and the lazy definition. We also checked the 18 move chain on the 40,320-state projection. We also checked the orientation model by verifying that the support of the distribution grew to 88,179,840 states.

2.6 Sensitivity analyses

Additional analyses of the 40,320-state random walk were performed. The "laziness" parameter p_{stay} was set to 0.15, 0.25, 0.5, 0.75 and 0.9, and the TV-distance allowed decay to 0.25 in each case. The three generating sets of moves used for these additional analyses were: the uniform 12-moves set; the 8-moves set obtained by removing one opposite face pair (L and R) from the 12-moves set, and verified by BFS to be able to reach all 40,320 states; and the 12-moves set with the four U and D moves given twice the weight of the other eight moves. The corresponding Shannon entropy was also tracked in all cases.

$$H\left(P_t\right) = -\sum_s P_t(s) log_2 P_t(s), \quad 0 \leq H\left(P_t\right) \leq log_2 N \tag{5}$$

where N = 40,320. Entropy is not an independent metric, since it is a function of the same P_t. It is linked to TV distance through the Kullback–Leibler divergence from uniform and Pinsker's inequality (Cover & Thomas, 2006):

$$D\left(P_t \| \pi\right) = log_2 N - H\left(P_t\right), \quad TV\left(P_t, \pi\right) \leq \sqrt{\frac{D\left(P_t \| \pi\right) ln\, 2}{2}} \tag{6}$$

Finally, probability mass was grouped by each state's BFS distance from solved, to show how the distribution spreads out over time.

All computations were written in Python 3.14 with NumPy 2.5 and SciPy 1.18 (sparse eigensolver), in double precision, on a laptop with an Apple M5 processor and 16 GB of memory. Every result for the 40,320-state walk takes about 5 seconds in total. Each run of the full 88,179,840-state chain takes about 2.5 minutes for 30 steps and uses about 2.5 GB of memory.

3. RESULTS

3.1 Mixing time of the corner-permutation walk

The figure plots the exact k-move walk in Figure 1a against the correct parity class. The TV distance is 0.274 after 9 quarter turns and 0.231 after 10. Hence the mixing time is 10 quarter turns. The lazy chain's mixing time at $p_{stay} = 0.5$ is the first time that the TV distance falls below 0.25, which is at tick 20 (TV distance = 0.248) / 0.270 at tick 19. In terms of expected moves this is 10.0. The continuous-time chain mixes at 10.4 expected moves. The plot against expected moves is very similar to the plot of the exact k-move walk in Figure 1a. The exact k-move walk is also plotted against the uniform distribution on all of S8. As predicted by the parity argument, it never falls below 0.5, e.g. 0.511 after 10 moves. In the rest of this paper we will plot against the lazy chain's mixing time at $p_{stay} = 0.5$ and use this to also give the mixing time of the other two versions at several different thresholds in Table 1.

Table 1. Mixing time of the corner-permutation walk (12 quarter-turn generators) at four thresholds. The lazy chain uses $p_{stay} = 0.5$. Expected moves = ticks × 0.5 for the lazy chain and the Poisson time for the continuous-time chain.

Threshold ε	Exact *k*-move walk (vs. parity class)	Lazy chain (ticks)	Lazy chain (expected moves)	Continuous time (expected moves)
0.5	6	12	6.0	6.0
0.25	10	20	10.0	10.4
0.1	15	31	15.5	16.0
0.01	28	57	28.5	29.7

1

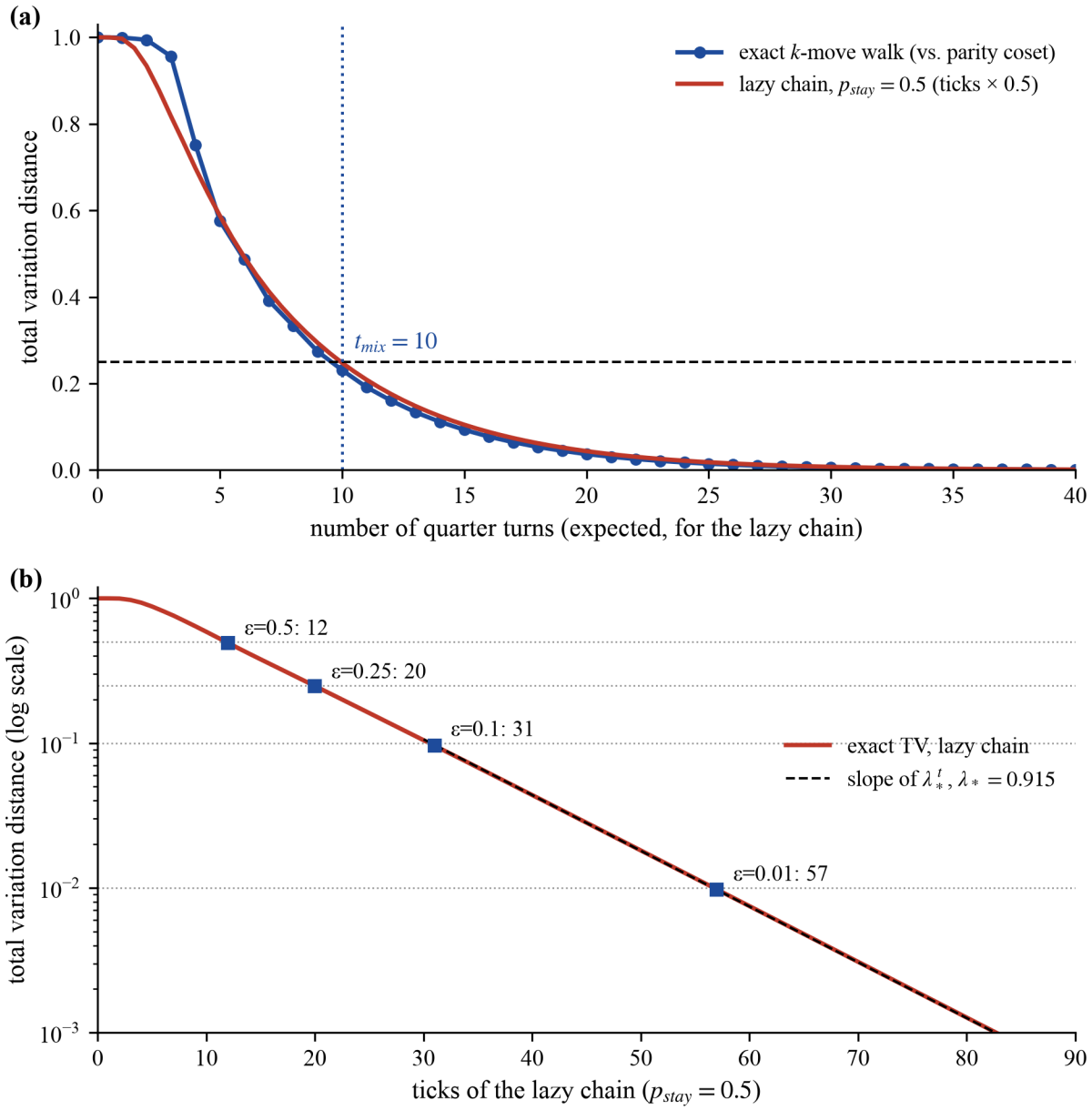


Figure 1. (a) Exact TV distance of the corner-permutation walk. Blue: the exact *k*-move walk, measured against the uniform distribution on the permutations of the correct parity. Red: the lazy chain (p_{stay} = 0.5), plotted against expected moves (ticks × 0.5). The dashed line is ε = 0.25. (b) The lazy chain's TV distance on a logarithmic scale against ticks, with the mixing time at ε = 0.5, 0.25, 0.1 and 0.01 marked. The dashed line has the slope of λ_*^t with λ_* = 0.915, the chain's second-largest eigenvalue.

3.2 Shape of the curve and the spectral gap

The lazy chain's TV distance falls to 0.75 at tick 8, to 0.5 at tick 12, to 0.1 at tick 31 and to 0.01 at tick 57 (Figure 1b). The window from $\varepsilon = 0.75$ to $\varepsilon = 0.25$ is 12 ticks long, and the window from $\varepsilon = 0.5$ to $\varepsilon = 0.1$ is 19 ticks long. Both are about as long as the mixing time itself, so the data show no abrupt drop. Instead, Figure 1b shows the shape stated in the hypothesis: a short plateau, followed by a straight line on the logarithmic scale, which means geometric decay.

The second-largest eigenvalue of the non-lazy 12-move matrix is $\lambda_2 = 0.830$, with multiplicity greater than one; the matrix also has the eigenvalue -1. For the lazy chain this gives $\lambda_* = 0.915$ and a relaxation time of $t_{\text{rel}} = 11.8$ ticks. Between ticks 40 and 57 the TV distance falls by a factor of 0.9155 per tick, which matches $\lambda_* = 0.9151$ (Figure 1b, dashed line). The bounds in Eq. 4 give $8 \le t_{\text{mix}}(0.25) \le 141$ ticks. The observed value of 20 is much closer to the lower bound, because the upper bound allows for the worst case over all starting distributions.

3.3 Laziness

Mixing time measured in ticks increases steeply with p_{stay}, from 12 ticks at $p_{\text{stay}} = 0.15$ to 103 ticks at $p_{\text{stay}} = 0.9$. In expected moves, however, it stays between 9.75 and 10.30 (Table 2 and Figure 2). When the curves are plotted against expected moves (Figure 2b), they almost lie on top of one another.

Table 2. Mixing time ($\varepsilon = 0.25$) of the lazy corner-permutation chain for five values of p_{stay}.

p_{stay}	Mixing time (ticks)	Expected moves
0.15	12	10.20
0.25	13	9.75
0.50	20	10.00
0.75	41	10.25
0.90	103	10.30

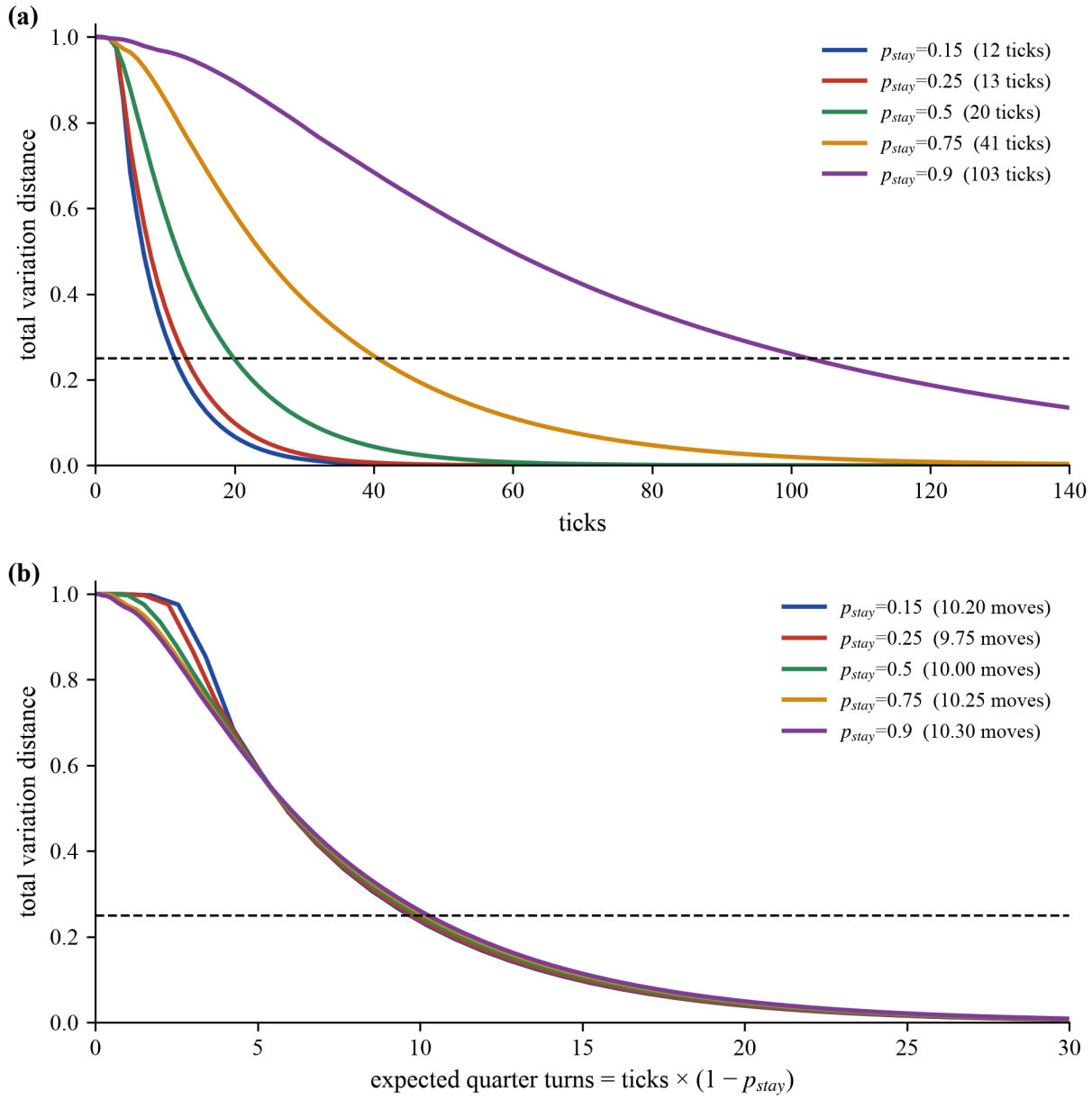


Figure 2. Exact TV distance of the lazy corner-permutation chain for five values of p_{stay}, (a) against ticks and (b) against expected moves, ticks × (1 − p_{stay}). The dashed line is ε = 0.25.

3.4 Generating set

The 8 moves (leaving out L and R) reach all 40,320 permutations in 9 moves with a larger diameter. This setup mixes in 31 ticks (10.0+2.5*20=15.5 expected moves) or 55% more than the 12 to 20 moves in the previous setup. Symmetrically, F and B, or U and D, turns would result in the same length and mixing time of 31 moves. Doubling the weight of the U and D moves while keeping the other 12 would mix in 22

ticks (11.0 expected moves, +10%). The spectral gap into 8-moves and weighted 12-moves (Table 3 and Figure 3) predict these two results. The lazy relaxation time is 11.8 ticks for the 12 uniform moves, 17.7 for the 8-moves (1.50*11.8=17.7) and 12.7 for the weighted 12-moves (1.08*11.8=12.7). These numbers are close to the ratios of the mixing times, 1.55 and 1.10.

Table 3. Effect of the generating set on the lazy chain (p_{stay} = 0.5, ε = 0.25). λ_2 is the second-largest eigenvalue of the non-lazy transition matrix; t_{rel} is the lazy chain's relaxation time. Removing F/B or U/D instead of L/R gives identical values.

Generating set	States reached	Diameter	λ_2	t_{rel} (ticks)	Mixing time (ticks)	Expected moves
12 moves, uniform	40,320	8	0.830	11.8	20	10.0
8 moves (no L, R)	40,320	9	0.887	17.7	31	15.5
12 moves, U/D double weight	40,320	8	0.842	12.7	22	11.0

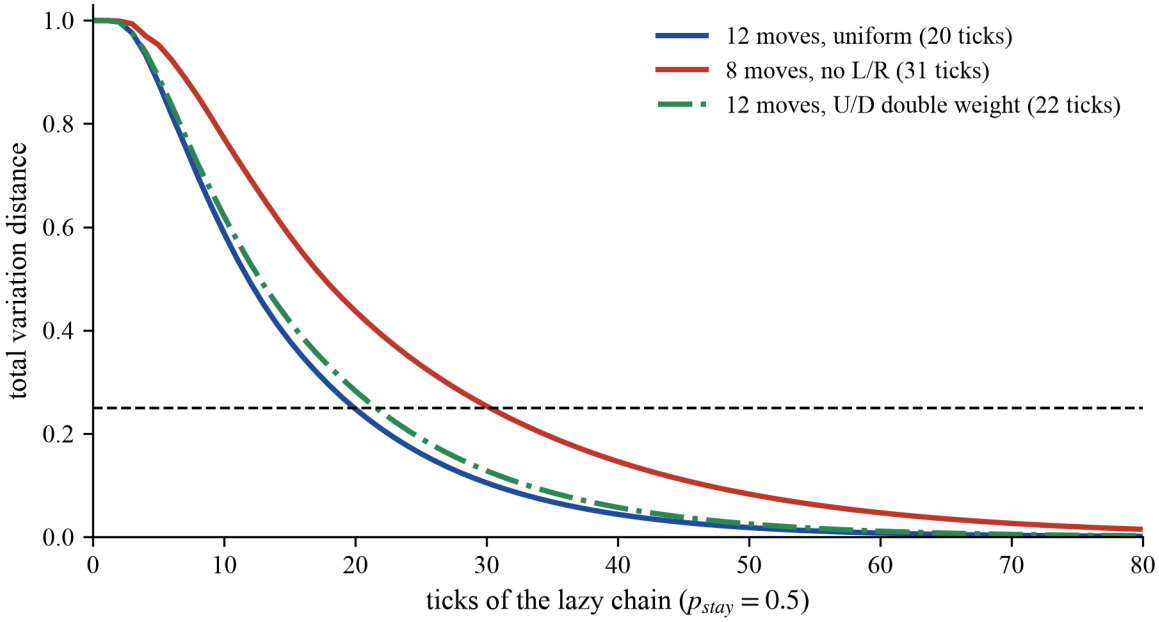


Figure 3. Exact TV distance of the lazy corner-permutation chain (p_{stay} = 0.5) for three generating sets. The dashed line is ε = 0.25.

3.5 Starting state, entropy and the distance profile

As a check on our code, starting the random walk from a scrambled cube position (U,F,L from a solved cube) gives us the same TV diagram as if we had started from a solved cube position. Since a random walk on a group is translated by a group element and the distribution is shifted by that same group element, the uniform distribution is invariant under such a shift. This is a theorem of Levin et al. (2017). The two plots above are thus a check on our code, not a new and interesting property of the cube.

Entropy gets to 90% of $\log_2 N = 15.299$ bits at tick 12. TV distance to get to 0.25 takes 20 steps (Fig. 4a). Note that the number of steps for 90% entropy and for 25% TV distance are not a priori expected to be close. Instead, we use the Kullback–Leibler divergence D to get a precise link between them. At tick 12, D = 1.38 bits. Therefore, by Pinsker's inequality, TV distance is bounded by 0.69 (in reality it is 0.49). At tick 20, D = 0.28 bits. Therefore, the bound from Pinsker's inequality is 0.31 (in reality TV distance is 0.25). The Pinsker bound in Fig. 4b always stays above the curve that corresponds to the current state of the walk, but gets close in the tail.

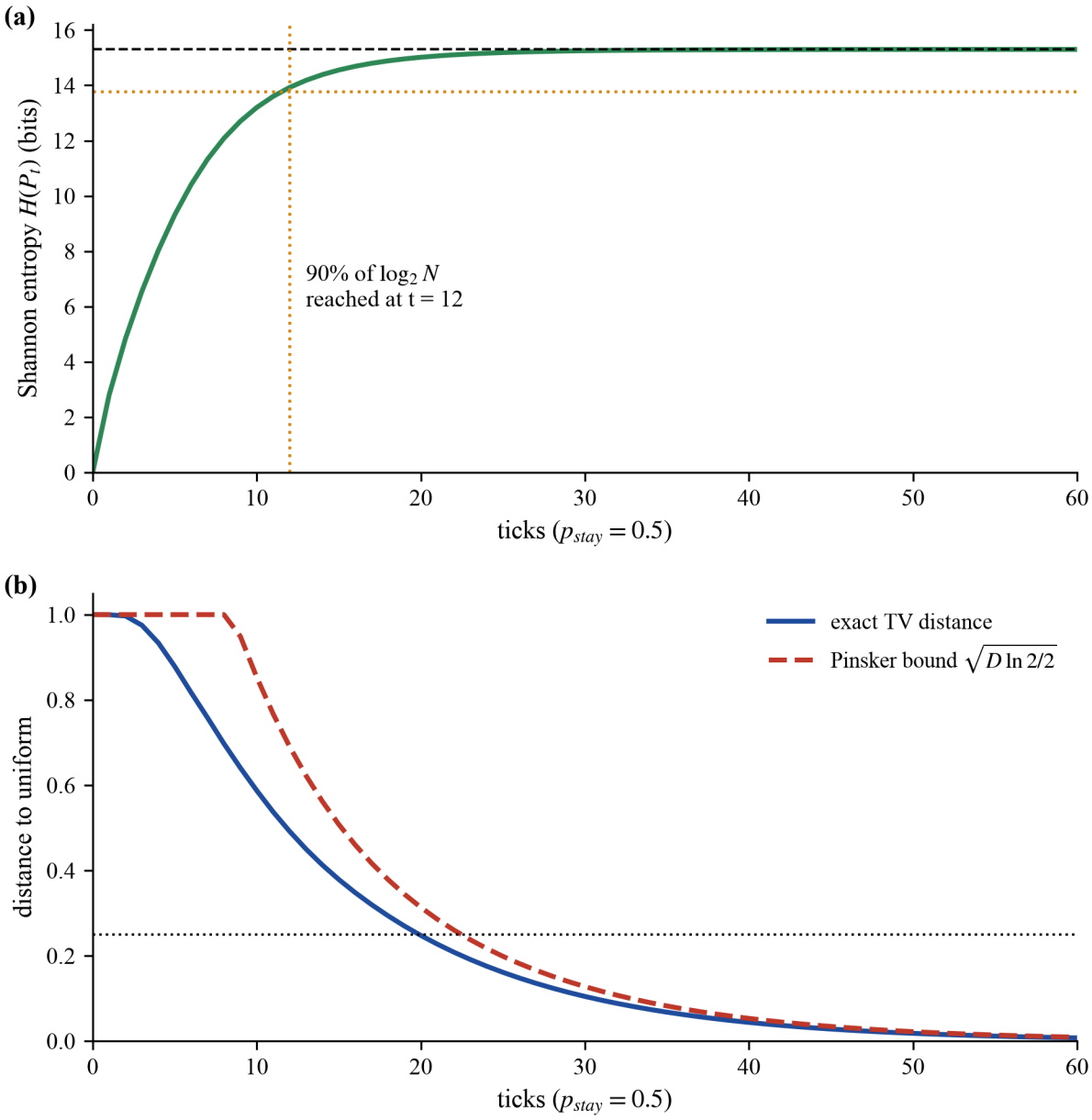


Figure 4. (a) Shannon entropy of the lazy corner-permutation chain ($p_{stay} = 0.5$). Maximum Shannon entropy is $\log_2 40{,}320 = 15.299$ bits. The dotted lines represent 90% of the maximum Shannon entropy. It reaches 90% maximum Shannon entropy at tick 12. (b) TV distance (solid line) and its upper bound Pinsker bound (dashed line) computed from the Kullback–Leibler divergence $D = \log_2 N - H$.

Figure 5 illustrates how the probability mass is aggregated by the distance to be solved for the target distribution. In marked contrast to a flat line, a hump-shaped line is found with the bulk of the states' probability mass at distances of 5 to 7 from the solved cases. The distribution's TV distance decreases when the distribution is grouped (for example, 0.181 as opposed to 0.248 at tick 20; 0.542 as opposed to 0.587 at tick 10). Both values are graphically shown in all the panels.

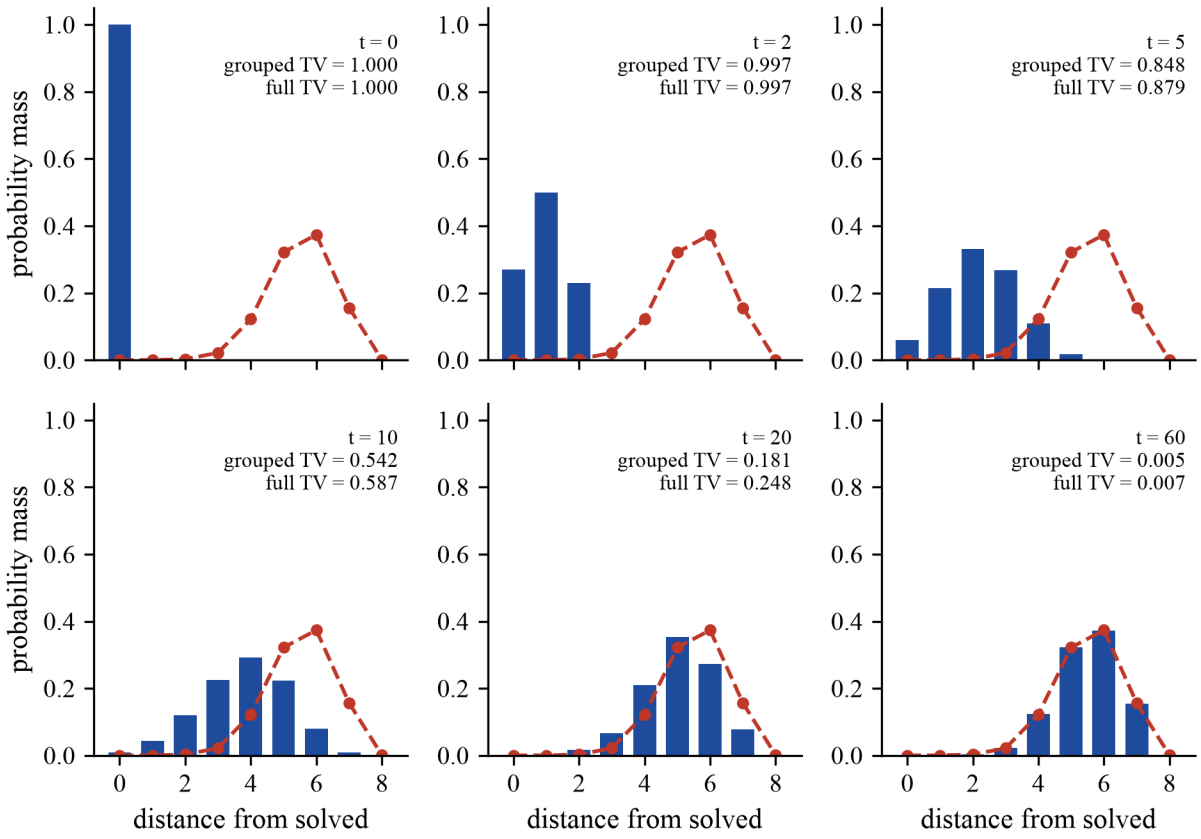


Figure 5: Probability mass of the lazy corner-permutation chain (p_{stay} = 0.5) binned by distance from the solved position. The red dots correspond to the probability mass of the uniform distribution on the same set of states binned the same way. Each panel reports the TV distance of the binned distribution as well as the full TV distance from the solved position for all 40,320 states in the sample. As always, the grouped value is the smaller of the two.

3.6 Comparison with the full 2×2×2 cube

Mixing time of the corner-permutation projection under the 18-move chain is 0.273 after 10 moves and 0.233 after 11 moves (totaling 11 moves in total), and the exact computation of the mixing time of the full fixed-frame corner group is 0.281 after 18 moves and 0.242 after 19 moves (totaling 19 moves in total) [4, 5]. This value agrees with the value reported in [4, 5] for both the orientation model and the method. The number of states that the distribution covered after 11 moves was 88,179,840 states. The TV distance for the full group of 12 quarter turns needed 22 moves according to the exact k-move definition (TV = 0.261

after 21 moves and TV = 0.229 after 22 moves), and 44 ticks for the lazy chain (22.0 expected moves). The distribution for the permutation of 10 moves covered all states after 10 moves (Table 4 and Figure 6). Orientation adds 8 moves on average under the 18-move chain, and 12 quarter turns (12-move chain) adds 12 moves on average.

Table 4. Exact mixing time (ε = 0.25) of the corner-permutation projection and of the full fixed-frame 2×2×2 corner group.

Chain	**Corner permutation (40,320 states)**	**Full 2×2×2, fixed frame (88,179,840 states)**	**Added by orientation**
18 moves (quarter, half, three-quarter turn)	11	19	8
12 quarter turns, exact *k*-move (vs. parity class)	10	22	12
12 quarter turns, lazy ($p_{stay} = 0.5$)	20 ticks (10.0 moves)	44 ticks (22.0 moves)	24 ticks (12.0 moves)

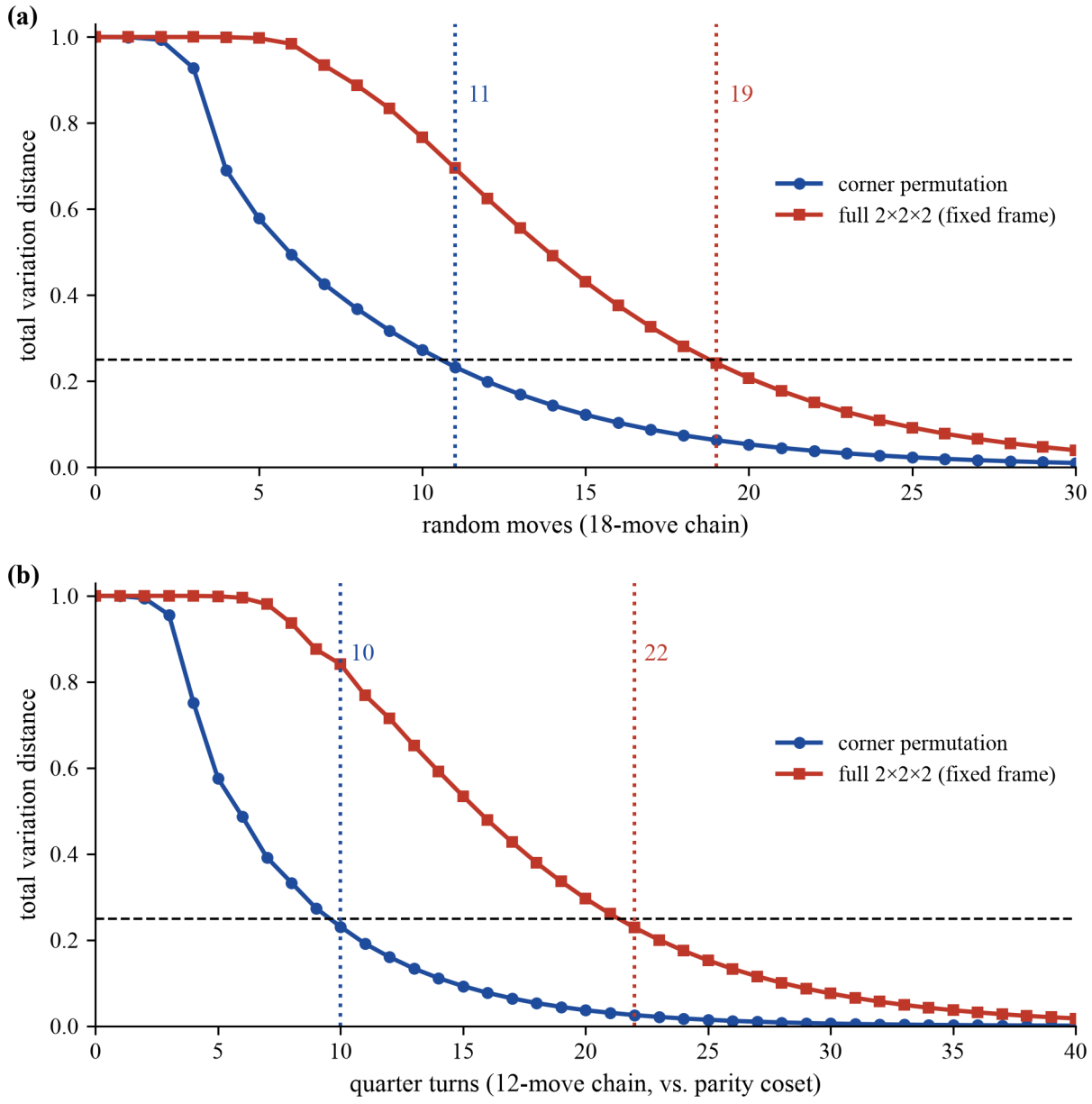


Figure 6. Exact TV distance of the corner-permutation projection (blue; 40,320 states) and of the full 2×2×2 corner group in a fixed frame (red; 88,179,840 states). (a) The 18-move chain used by Qu et al. (2024) and Fernique (2025). (b) The 12 quarter-turn chain, measured against the uniform distribution on the correct parity class. The dashed line is $\varepsilon = 0.25$.

4. DISCUSSION

Ten quarter turns suffice to solve the corners of the 2×2×2 puzzle according to three definitions of move count: the exact k-move walk (10 moves), the lazy chain (20 ticks, or 10.0 expected moves) and the continuous-time chain (10.4). The answer to the research question is the number of quarter turns of these 10 moves, i.e., 10 quarter turns. For the parity of a move count k, all k moves form one parity class, and thus the only reasonable target distribution is the uniform distribution on that class of parity. The quarter-turn walk in the textbook sense does not mix for the full uniform distribution on all moves.

We interpret the laziness sweep in Figure 2(b) as a consistency check. We can run the underlying walk on a random clock, i.e., a Binomial(t, $1-p_{stay}$) distribution of real moves after t ticks. Therefore tmix($1-p_{stay}$) is constant and close to 10 for all values of p_{stay}, as shown in Figure 2(b). The small variations come from the spread of the distribution of the number of real moves in the lazy chain, and they matter most for a few ticks. Instead of showing independently that 10 quarter turns are robust, the lazy construction just transforms time, and the extended window for p_{stay}=0.9 shows that for this value also, a window of 103 ticks is sufficient to reach a TV of 0.1 with an expected number of moves of 10.3.

The observed shape of the curve (a short plateau followed by a steep decline of a geometric decay of rate 0.9155 per tick) is as expected for a chain of this size. In particular, there is no sharp cutoff as one might expect for a chain of this size. Rather, the window in which the transition variable TV attains values greater than 0.5 to 0.1 is roughly of the same length as the mixing time itself. This is not evidence for or against the hypothesis of a given cutoff for this chain, since a family of chains would be more appropriate to test this hypothesis. Corner-permutation walks on puzzles with increasing numbers of pieces would be a natural family to test.

We also use the spectral gap to explain the behavior of the generating set. Removing a face pair from the generating set decreases the spectral gap by about one third. The relaxation time, however, increases by about 50%, and the mixing time increases by about 55% in this case. This difference to 55% is due to the exact computation of mixing time. Re-weighting the same twelve moves results in only a minor change in the spectral gap, i.e. the generator effect is not only caused by the increased diameter of 9 versus 8. Also, fewer generators result in less effective moves to spread mass, and the spectral gap of the generator set is a direct measure for this.

The comparison with the full cube (Section 3.6) is probably the result with the largest number of implications. While a chain on the corners of the full cube mixes in 11 moves, the full cube itself mixes in 19 moves. So about11 moves of the 19 on the full cube are spent to mix the corners, and about 8 moves to mix the respective orientations, while under quarter turns on the 2×2×2 puzzle the corners and orientations mix in 10 and 12 moves respectively. Thus the work here, while being different from the work in [34, 41] (where samples are used to estimate the distance from solved statistic and a projection onto this statistic is used), provides exact results for the partition into permutations and orientations (while [34] only provides a lower bound for the number of quarter turns to mix the corners, namely 11, which is different from the lower bound for the full cube of 19 moves, where 11 moves would mix the corners,

while the remaining 8 moves would mix the respective orientations, see Section 3.6 for details). It also shows that the permutation projection gives a valid but loose lower bound for the full cube, namely 11 instead of 19 moves. The curves for the two values and their decomposition into permutations and orientations are provided in Section 3.6.

However, our main results, for example the number of quarter turns needed to mix a corner of a puzzle, are dependent on the generating set we chose for the group. The results also only consider the permutation of corners, not the whole puzzle. We use the same spatial frame for our full cube as Qu et al. and Fernique, thus the number of states for the positions up to whole-cube rotation is 3,674,160. The mixing time of this walk can only be equal or smaller than the mixing time of the walk considered in this paper. Last but not least, a finite chain, like the ones considered in this paper, cannot be used to test for cutoff. Natural extensions of this work would be the rotation-reduced 2×2×2 walk, the 3×3×3 corners (which are the same as the 2×2×2 corners considered in this paper) and a family of puzzles where cutoff could be tested for.

5. CONCLUSION

The corner permutation of the 2×2×2 Rubik's cube is mixed to TV distance 1/4 after 10 random quarter turns. The k-move walk, lazy chain and continuous-time chain all compute this quantity exactly (to within a parity class). The TV curve has a short plateau followed by a curve decreasing at the rate of the spectral gap. The spectral gap for the reduced cube also explains the 55% slowdown when removing one face pair. For the full corner group, the exact computation of the mixing time of the 18-move chain is 19. It shows that the corner permutation mixes about 11 moves of corner permutation and 8 moves of orientation. Note that while exact computation on small projections is of interest for the 3×3×3 problem, it will not tell us which parts of the cube take the longest to become random.

CODE AVAILABILITY

All code used to produce the results, tables and figures in this paper is publicly available at https://github.com/JaiGulyani/Rubik-s-2x2x2-Mixing-Time.git.The repository contains the move tables, the BFS, the exact iteration for both the 40,320-state and 88,179,840-state chains, the spectral computation and the figure scripts. Running it reproduces every reported number exactly.

REFERENCES

Aldous, D. (1983). Random walks on finite groups and rapidly mixing Markov chains. In J. Azéma & M. Yor (Eds.), *Séminaire de Probabilités XVII* (Lecture Notes in Mathematics, Vol. 986, pp. 243–297). Springer.

Aldous, D., & Diaconis, P. (1986). Shuffling cards and stopping times. *The American Mathematical Monthly, 93*(5), 333–348.

Bayer, D., & Diaconis, P. (1992). Trailing the dovetail shuffle to its lair. *The Annals of Applied Probability, 2*(2), 294–313.

Chen, J. (2004). *Group theory and the Rubik's Cube* [Lecture notes]. Harvard University. Retrieved September 28, 2026, from https://people.math.harvard.edu/~jjchen/docs/Group%20Theory%20and%20the%20Rubik's%20Cube.pdf

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). Wiley.

Diaconis, P. (1988). *Group representations in probability and statistics* (IMS Lecture Notes—Monograph Series, Vol. 11). Institute of Mathematical Statistics.

Diaconis, P. (1996). The cutoff phenomenon in finite Markov chains. *Proceedings of the National Academy of Sciences, 93*(4), 1659–1664.

Diaconis, P., & Shahshahani, M. (1981). Generating a random permutation with random transpositions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 57*(2), 159–179.

Fernique, T. (2025). *How not to scramble a Rubik's cube* (arXiv:2509.12134). arXiv.

Hirata, S. (2024). *Probabilistic estimates of the diameters of the Rubik's Cube groups* (arXiv:2404.07337). arXiv.

Joyner, D. (2008). *Adventures in group theory: Rubik's Cube, Merlin's Machine, and other mathematical toys* (2nd ed.). Johns Hopkins University Press.

Kunkle, D., & Cooperman, G. (2007). Twenty-six moves suffice for Rubik's Cube. In *Proceedings of the 2007 International Symposium on Symbolic and Algebraic Computation (ISSAC '07)* (pp. 235–242). ACM.

Levin, D. A., Peres, Y., & Wilmer, E. L. (2017). *Markov chains and mixing times* (2nd ed.). American Mathematical Society.

Qu, Y., Rokicki, T., & Yang, H. (2024). *Rubik's Cube scrambling requires at least 26 random moves* (arXiv:2410.20630). arXiv.

Rokicki, T. (2010). Twenty-two moves suffice for Rubik's Cube. *The Mathematical Intelligencer, 32*(1), 33–40.

Rokicki, T. (2014). Towards God's number for Rubik's Cube in the quarter-turn metric. *The College Mathematics Journal, 45*(4), 242–253.

Rokicki, T., & Davidson, M. (2014). *God's number is 26 in the quarter-turn metric*. Retrieved September 28, 2026, from https://cube20.org/qtm/

Rokicki, T., Kociemba, H., Davidson, M., & Dethridge, J. (2014). The diameter of the Rubik's Cube group is twenty. *SIAM Review, 56*(4), 645–670.

Saloff-Coste, L. (2004). Random walks on finite groups. In H. Kesten (Ed.), *Probability on discrete structures* (Encyclopaedia of Mathematical Sciences, Vol. 110, pp. 263–346). Springer.

Singmaster, D. (1981). *Notes on Rubik's Magic Cube*. Enslow Publishers.