Information content in the Nagel-Schreckenberg cellular automaton traffic model

M. Blue*
Systems Engineering and Integration Group, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

B. W. Bush†
Energy and Environmental Analysis Group, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

(Received 29 August 2001; revised manuscript received 22 November 2002; published 25 April 2003)

We estimate the set dimension and find bounds for the set entropy of a cellular automaton model for single lane traffic. Set dimension and set entropy, which are measures of the information content per cell, are related to the fractal nature of the automaton [S. Wolfram, Physica D **10**, 1 (1989); *Theory and Application of Cellular Automata*, edited by S. Wolfram (World Scientific, Philadelphia, 1986)] and have practical implications for data compression. For models with maximum speed $v_{\max}$, the set dimension is approximately $\log_{(v_{\max}+2)}2.5$, which is close to one bit per cell regardless of the maximum speed. For a typical maximum speed of five cells per time step, the dimension is approximately 0.47.

DOI: 10.1103/PhysRevE.67.047103 PACS number(s): 89.40.−a, 05.40.−a, 89.75.Kd

I. INTRODUCTION

In [1] Nagel and Schreckenberg describe a cellular automaton model for traffic flow that resembles traffic in real life [2,3]. In particular, the TRANSIMS project [4] has adapted this model to simulate traffic in entire cities. Recently, TRANSIMS has focused on modeling the city of Portland, Oregon, but due to the size of Portland, this simulation generates a large amount of data (approximately one terabyte of state evolution information for a 24-hour simulation), making data compression a necessity. Understanding the possible allowed configurations of the automaton (i.e., the set entropy) furthers the construction of efficient data compression schemes.

Claude Shannon introduced information theory as a way to describe the information content in a given system [5]. Set entropy measures the number of possible states occurring in a cellular automaton; set dimension is the limit of the set entropy as the spatial extent approaches infinity [6,7]. To measure this quantity for cellular automata, suppose there are N letters in our alphabet. Then for a cellular automaton of dimension k (the number of lanes on the roadway, in the automaton considered here) and size n (the length of the roadway), there are N^{nk} possible states for the cellular automaton. Let $|{}_kA_n|$ be the number of states that may actually occur in the automaton, given the rules constraining its evolution in time. The set entropy $s(k,n)$ is given by

$$s(k,n)=\frac{\log_N|{}_kA_n|}{nk}=\frac{\ln|{}_kA_n|}{\ln N^{nk}}$$

and the set dimension d is given by

$$d=\lim_{n\to\infty}s(k,n)=\lim_{n\to\infty}\frac{\log_N|{}_kA_n|}{nk}=\lim_{n\to\infty}\frac{\ln|{}_kA_n|}{\ln N^{nk}}, \qquad (1)$$

refer to Refs. [5,6,8] for more details.

In this Brief Report, we construct bounds for the set entropy and estimate the set dimension for a single-lane-roadway ($k=1$) model [1,3,4].

II. DESCRIPTION OF THE MODEL

In the single-lane model, the system is a grid of n sites (or cells), $c_0c_1\cdots c_{n-1}$. Each cell c_i can either be empty, contain a stopped vehicle (with speed 0), or contain a moving vehicle with speed v. For a vehicle x, let c_x be its position, and v_x be its velocity. This information is encoded in the cellular automaton by the number v_x in cell c_x. Let Δ (the "gap") denote the amount of space between vehicle x and the vehicle directly in front of x. The following three rules are used to determine the velocity of each car at the next time step [1,3,4].

Rule 1. Accelerate if you can: if $v_x<\Delta$ and $v_x<v_{\max}$, then $v_x:=v_x+1$.

Rule 2. Decelerate to avoid rear-end collisions: if $v_x>\Delta$, then $v_x:=\Delta$.

Rule 3. Stochastic behavior: if $v_x>0$ and $r<p_{\rm d}$, then $v_x:=v_x-1$.

In Rule 3, $p_{\rm d}$ is the probability that a vehicle will slow down for no reason, and r is a uniform random variable. The rules can be modified to allow for individual cars to have different preferred speeds, but for our purposes a universal maximum speed $v_{\max}$ and deceleration probability $p_{\rm d}$ suffice. To change states at a given time step, first determine the velocity of each vehicle x, then move x from cell c_x to cell c_{x+v_x}. Thus, if the pair of cells labeled "5,0" occurs in the automaton, the 5 and the 0 represent the speed of the vehicles based upon the previous state.

III. BOUNDS ON ENTROPY

Let A_n denote the set of one-lane blocks of n cells that may occur in an automaton. We refer to these as *allowed blocks*. We are interested in computing the size $|A_n|$ of A_n for any n. However, finding a closed form for $|A_n|$ proves to be

*Electronic address: mblue@lanl.gov
†Electronic address: bwb@lanl.gov

very difficult, even after discovering a successive listing for the blocks that can never occur in the automaton (see Ref. [8] for the listing). Thus, in order to compute the dimension, we find upper and lower bounds for $|A_n|$, which tend to similar limits as $n \to \infty$. (The difference between the limits turns out to be approximately 0.13.) Other work on allowed states in traffic models has been performed by Schadschneider and Schreckenberg [9], who identify the "garden of Eden" states for the cases $v_{\max}=1$ and $v_{\max}=2$.

We begin by making two observations about the driving rules. First, suppose at time t there is a moving car with speed v in cell c_i. This implies that at time $t-1$ the vehicle was in cell c_{i-v}. In order for the vehicle to move at speed v, the cells $c_{i-v}, c_{i-v+1}, \ldots, c_{i-1}$ must have been empty. Furthermore, these cells must remain empty at time t, because any vehicle behind the car will never move forward more cells than the gap between the two cars at time $t-1$. For the second observation we again suppose at time t there is a car x with speed v_x in cell c_i, but now we suppose that there is a car y in cell c_{i+1} with speed v_y. We claim that $v_y=0$. This follows because speed $v_x \leqslant \Delta$. Thus, vehicle x can move no further than the cell behind vehicle y at time $t-1$. If vehicle y were moving there would be a gap between vehicles x and y at time $t-1$.

Suppose $B=c_0 c_1 \cdots c_{n-1}$ is an allowed block of size n. Our upper and lower bounds are based on the above observations illustrated for block B below.

(i) If there is a car in cell c_m, $0 \leqslant m < n-1$, then cell c_{m+1} is either empty or contains a vehicle with speed 0. In any case, there are only two choices for cell c_{m+1}, provided there is a car in cell c_m.

(ii) If there is a car in cell c_m with speed v, then all cells $c_{m-1}, c_{m-2}, \ldots, c_{m-v}$ must be empty.

We wish to find an upper and lower bound for $|A_{n+1}|$ in terms of $|A_n|$. Let $A_n(\bigcirc)$ be the set of allowed blocks of length $n+1$ that end with an empty cell. Similarly, let $A_n(s)$ be the set of blocks of length $n+1$ where the last cell contains a vehicle with speed s. We first note that

$$A_{n+1}=A_n(\bigcirc) \bigcup_{s=0}^{v_{\max}} A_n(s),$$

which yields

$$|A_{n+1}|=|A_n(\bigcirc)|+\sum_{s=0}^{v_{\max}} |A_n(s)|.$$

Since an empty cell may be preceded by an empty cell or a vehicle of any speed, $|A_n(\bigcirc)|=|A_n|$. Similarly, a vehicle of speed 0 may also be preceded by an empty cell or a vehicle of any speed; hence, $|A_n(0)|=|A_n|$.

Let us now consider an allowed block of length $n+1$, which has a 1 in cell c_n (recall our blocks begin with cell c_0). Cell c_{n-1} must be empty, but cell c_{n-2} may be empty or contain a car of any speed. Thus, $|A_n(1)|=|A_{n-1}|$.

Turning to $A_n(2)$, we consider the allowed blocks of length $n+1$ where cell c_n contains a 2. In this case both cells c_{n-1} and c_{n-2} must be empty. Cells c_{n-3} is permitted to be either empty or contain a vehicle with speed greater than 0. A car of speed 0 in cell c_{n-3} is a violation of observation 1, as the state "0,○,○,2" can only arise from forbidden states "0,2" or "0,1" in the previous time step (see Refs. [8] and [9]). Thus, $|A_n(2)|=|A_{n-2}|-|A_{n-3}|$.

For $A_n(s)$ with $s>2$, the computation is much more complicated. However, we note that cells $c_{n-1}, \ldots, c_{n-s}$ all must be empty. As with the $s=2$ case, cell c_{n-s-1} cannot contain a stopped vehicle. With speed $s>2$ there are more restrictions, but we will just note that in this case $|A_n(s)| \leqslant |A_{n-s}|-|A_{n-s-1}|$.

Let us first use the following simple lower bound:

$$|A_{n+1}| \geqslant |A_n(\bigcirc)|+\sum_{s=0}^{2} |A_n(s)|=|A_n|+|A_n|+|A_{n-1}| +(|A_{n-2}|-|A_{n-3}|),$$

but we can simplify this further for easier computation if we underestimate $|A_n(2)|$ by considering only those blocks with cells c_{n-1}, c_{n-2}, and cell c_{n-3} empty. This gives $|A_n(2)| \geqslant |A_{n-3}|$, so we now have

$$|A_{n+1}| \geqslant L_n \equiv 2|A_n|+|A_{n-1}|+|A_{n-3}| \quad \text{for } n \geqslant 4. \tag{2}$$

We now turn to the upper bound, using the above information:

$$\begin{aligned}|A_{n+1}|&=|A_n(\bigcirc)|+\sum_{s=0}^{v_{\max}} |A_n(s)| \leqslant |A_n|+|A_n|+|A_{n-1}|\\ &\quad +(|A_{n-2}|-|A_{n-3}|)+(|A_{n-3}|-|A_{n-4}|)+\cdots\\ &\quad +(|A_{n-v_{\max}}|-|A_{n-v_{\max}-1}|)=2|A_n|+|A_{n-1}|\\ &\quad +|A_{n-2}|-|A_{n-v_{\max}-1}| \leqslant 2|A_n|+|A_{n-1}|+|A_{n-2}|.\end{aligned}$$

This gives us an equation for the upper bound:

$$|A_{n+1}| \leqslant U_n \equiv 2|A_n|+|A_{n-1}|+|A_{n-2}| \quad \text{for } \quad n \geqslant 4. \tag{3}$$

To begin our entropy estimation, we first bound $|A_{n+1}|$ by simplifying the bounds in Eqs. (2) and (3):

$$2|A_n| < |A_{n+1}| < 4|A_n| \quad \text{for } \; n \geqslant 4. \tag{4}$$

We would now like to combine Eqs. (3) and (4) to get a better upper bound for $|A_{n+1}|$ of the form $\rho|A_n|$. (Bounds of the form $\rho|A_n|$ are important because they allow the calculation of entropy in the limit $n \to \infty$.) First notice that Eq. (4) implies that $|A_n| < \frac{1}{2}|A_{n+1}|$, for all n. We now compute

$$\begin{aligned}|A_{n+1}| &\leqslant 2|A_n|+|A_{n-1}|+|A_{n-2}| \leqslant 2|A_n|+\tfrac{1}{2}|A_n|+\tfrac{1}{2}|A_{n-1}|\\ &\leqslant \tfrac{5}{2}|A_n|+\tfrac{1}{4}|A_n|=\tfrac{11}{4}|A_n|.\end{aligned}$$

So we could now use $\frac{4}{11}|A_{n+1}| < |A_n|$, instead of $\frac{1}{4}|A_{n+1}| < |A_n|$, to get a better lower bound for $|A_{n+1}|$. In fact, we can recursively improve lower and upper bounds of the form $a_m^{-1}|A_n| < |A_{n+1}| < b_m^{-1}|A_n|$, where a_m and b_m are successive improvements on the bounds. We start with Eq. (4),

TABLE I. Comparison of the exact value of $|A_n|$ with the lower bound L_n, the upper bound U_n, the scalable bounds L_n^* and U_n^*, and the states S_n observed in a computer simulation; here $v_{\max}=5$.

n	$\lvert A_n\rvert$	L_n	U_n	L_n^*	U_n^*	$\lvert S_n\rvert$
1	7					7
2	19					19
3	49					49
4	126					126
5	320	308	320	309	324	320
6	811	785	815	783	824	811
7	2 045	1 991	2 068	1 985	2 088	2 045
8	5 145	5 027	5 221	5 004	5 267	5 144
9	12 930	12 655	13 146	12 589	13 252	12 920
10	32 474	31 816	33 050	31 637	33 304	32 423
11	81 529	79 923	83 023	79 456	83 644	80 813
12	204 651	200 677	208 462	199 482	209 997	200 443
13	⩾513 583	503 761	523 305	500 732	527 128	490 764

which yields $a_1=\frac{1}{2}$ and $b_1=\frac{1}{4}$. From Eqs. (2) and (3), respectively, we obtain the recursion relations

$$a_{m+1}=(2+b_m+b_m^3)^{-1},$$

$$b_{m+1}=(2+a_m+a_m^2)^{-1}.$$

We are interested in $\alpha=\lim_{m\to\infty}a_m$ and $\beta=\lim_{m\to\infty}b_m$. As a_m and b_m are both monotonic and bounded, both α and β exist. We hope of course that $\alpha=\beta$, because in this case the upper and lower bound coincide. We are reduced to solving the following two equations in two unknowns:

$$\alpha=(2+\beta+\beta^3)^{-1},$$

$$\beta=(2+\alpha+\alpha^2)^{-1},$$

which can be solved numerically to obtain the following estimates: $\alpha\doteq 0.408\,704$ and $\beta\doteq 0.388\,237$, yielding $\gamma\equiv\alpha^{-1}\doteq 2.446\,76$ and $\delta\equiv\beta^{-1}\doteq 2.575\,74$. Thus, we have obtained the simple bounds

$$L_n^*\equiv\gamma|A_n|<|A_{n+1}|<U_n^*\equiv\delta|A_n| \text{ for } n\geqslant 4. \qquad (5)$$

Table I compares our bounds on $|A_n|$ with an exact enumeration of the allowed cellular automaton states and with the states observed in a computer simulation of the cellular automaton.

IV. DIMENSION

With a maximum speed of $v_{\max}$, there are $N=v_{\max}+2$ possible symbols for each cell in the Nagel-Schreckenberg traffic model. Thus, the set dimension is given by Eq. (1):

$$d=\lim_{n\to\infty}\frac{\log_{(v_{\max}+2)}|A_n|}{n}.$$

For the upper bound we have

$$|A_n|<\delta|A_{n-1}|=\delta^2|A_{n-2}|=\cdots=\delta^{(n-1)}|A_1|$$
$$=\delta^{n-1}(v_{\max}+2),$$

so

$$d<\lim_{n\to\infty}\frac{\ln[\delta^{n-1}(v_{\max}+2)]}{n\ln(v_{\max}+2)}$$
$$=\lim_{n\to\infty}\left[\frac{(n-1)\ln\delta}{n\ln(v_{\max}+2)}+\frac{\ln(v_{\max}+2)}{n\ln(v_{\max}+2)}\right]$$
$$=\frac{\ln\delta}{\ln(v_{\max}+2)}.$$

The lower bound is computed similarly to obtain

$$\frac{\ln\gamma}{\ln(v_{\max}+2)}<d.$$

Numerically, we now have

$$\frac{\ln 2.446\,76}{\ln(v_{\max}+2)}<d<\frac{\ln 2.575\,74}{\ln(v_{\max}+2)}$$

or

$$d\approx\frac{\ln 2.5}{\ln(v_{\max}+2)}.$$

Entropy, in general, is a measure of the ‘‘information content’’ per site [5,6]. In this case, for n sufficiently large, set A of allowed blocks is estimated by $A\approx(N^n)^d$. In terms of bits, $B\equiv\log_2 A=dn\log_2 N$ for n sites. Let I represent the information content in terms of bits per cell. Computing this, we obtain

$$I=\frac{\log_2 A}{n}\approx d\log_2 N\approx\frac{\ln 2.5}{\ln N}\log_2 N=\log_2 2.5\doteq 1.32.$$

Thus, little more than one bit is needed to encode the information contained in a single cell.

V. CONCLUDING REMARKS

We have bounded the set entropy and estimated the set dimension for the cellular automaton used in the Nagel-Schreckenberg model for the single-lane roads.

In the multilane case, symmetric lane-changing models [4,10] are prevalent. Before changing lanes, a vehicle typically must examine both the gap ahead and the gap behind in the other lane to avoid collisions. In such models, the lane changing rules do not limit the states reachable by the dynamics, i.e., if the vehicles on a particular segment of roadway have no preference for changing lanes at a given time, then their dynamics at that time does not depend on the configuration of vehicles in other lanes, and any of the states of single-lane traffic is possible. Depending on the exact form of the lane-changing rules, however, more states than just the product of single-lane states may be possible. An example of this is when a vehicle changes lanes to move into a gap between two vehicles where a vehicle normally would not be present. Thus, $|L_n|^k$ can be used as a lower bound for the multilane case, but an upper bound for it cannot be derived from U_n. The number of possible multilane states in excess of $|U_n|^k$ is relatively small, so our single-lane estimate of the dimensions provides a rough approximation for the multilane dimension.

To obtain optimal results in data compression, other entropies are interesting. These involve the probability of a given state occurring instead of just the possibility of the state occurring as in set entropy. This allows one to encode the most likely states with the fewest number of bits. These probabilities are very difficult to compute however, as they are extremely dependent on the traffic density and the particular city street network. In some cases, however, it may be possible to analyze these probabilities using techniques such as Markov chains.

ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy and the U.S. Department of Transportation.

[1] K. Nagel and M. Schrekenberg, J. Phys. I **2**, 2221 (1992).
[2] K. Nagel, D.E. Wolf, P. Wagner, and P. Simon, Phys. Rev. E **58**, 1425 (1998).
[3] P.M. Simon and K. Nagel, Phys. Rev. E **58**, 1286 (1998).
[4] TRANSIMS (TRansportation ANalysis and SIMulation System), URL: http://transims.tsasa.lanl.gov
[5] R.M. Gray, *Entropy and Information Theory* (Springer-Verlag, New York, 1990).
[6] *Theory and Applications of Cellular Automata*, edited by S. Wolfram (World Scientific, Philadelphia, 1986).
[7] S. Wolfram, Physica D **10**, 1 (1984).
[8] M. Blue and B.W. Bush, Los Alamos National Laboratory Tech. Rep. No. LA-UR-01-4277, 2001 (unpublished), URL: http://transims.tsasa.lanl.gov/PDF_Files/LAURs/LAUR-01-4277.pdf
[9] A. Schadschneider and M. Schreckenberg, J. Phys. A **31**, L225 (1998).
[10] M. Rickert, K. Nagel, M. Schreckenberg, and A. Latour, Physica A **231**, 534 (1996).