Airplane boarding, polynuclear growth, disk I/O scheduling and space-time geometry

Eitan Bachmat[1], Daniel berend[1,2], Luba Sapir[3] and Steven Skiena[4]
[1] *Department of Computer Science, Ben-Gurion University, Beer-Sheva 84105, Israel*
[2] *Department of mathematics, Ben-Gurion University, Beer-Sheva 84105, Israel*
[3] *Department of Management and Industrial engineering, Ben-Gurion University, Beer-Sheva 84105, Israel*
[4] *Department of Computer science, SUNY at Stony Brook, Stony Brook NY, 11794, USA*

We show that several discrete processes including airplane boarding, optimal I/O scheduling in disk drives, patience sorting and others can be asymptotically modeled by two dimensional Lorentzian geometry. In the case of aiplane boarding, boarding time is given by the maximal proper time of worldlines in the model. Fluctuations in boarding time about the maximal proper time are related to random matrix models. We then show how such models can be used to explain why commonly practiced airline boarding policies are ineffective and even detrimental.

We show a wide range of discrete processes originating in several distinct disciplines can be provided with a common description interms of two dimensional space-time geometry. Table 1 which will be explained below provides a dictionary for translating notions between the processes. This implies that the seemingly unrelated processes form a universality class which shares some common statistical properties, strongly related to random matrix theory (RMT) and that results which are obtained in one domain can be applied to the others. A particularly appealing example is provided by airplane boarding, a process which is personally experienced, daily, by millions of passengers worldwide. Airlines have developed various strategies in the hope of shortening boarding time, typically leading to announcements of the form " passengers from rows 40 and above are now welcome to board the airplane", often heard around airport terminals. We will later show that several other known processes are in fact special cases of airplane boarding. Therefore for the purpose of presentation we will consider our results in the context of this process which, rather surprisingly, has not been analyzed previously.

We describe the boarding process as follows: Passengers $1, ..., N$ are represented by coordinates $X_i = (q_i, r_i)$, where q_i is the index of the passenger along the boarding queue (1st, 2nd, 3rd and so on), and r is his/her assigned row number. We assume that the main cause of delay in airplane boarding is the time it takes passengers to organize their luggage and seat themselves, once they have arrived at their assigned row. The input parameters for our model are: u - The average amount of aisle length occupied by a passenger, luggage included.

w - The distance between successive rows. For the purposes of presentation we will assume that w is fixed.

b - Number of passengers per row.

D - Amount of time (delay) it takes a passenger to clear the aisle, once he has arrived at his designated row. We shall assume at first that D is fixed.

$p(q,r)$ - The joint distribution of a passenger's row and queue joining time. $p(q,r)$ is directly affected by the airline policy and the way passengers react to the policy.

We rescale (q,r) to [0,1]x[0,1]. and define on the unit square the Lorentzian metric

$$d\tau^2 = 4D^2 p(q,r)(dqdr + k\alpha(q,r)dq^2)$$

where $k = bu/w$ and $\alpha(q,r) = \int_r^1 p(q,z)dz$. In the absence of a boarding policy, $p(q,r) = 1$ and $\alpha = 1 - r$.

We define $T(X)$ as the maximal proper time (integral over $d\tau$) of a time-like trajectory ending at X. We also define $L(\tau)$ as the length (integral over $\sqrt{-d\tau^2}$) of the space-like curve which is defined by the equation $T(X) = \tau$.

Our statements are (A) The boarding time of passenger i is approximately $\sqrt{N}T(X_i)$

(B) The number dN of passenger with boarding time between $\sqrt{N}\tau$ and $\sqrt{N}(\tau + d\tau)$ is approximately $\frac{1}{2D}\sqrt{N}L(\tau)$

In the statements we use the word approximately to mean, ratio tending to 1 with probability approaching 1 as the number of passengers N tends to infinity.

Airplane boarding and Lorentzian geometry are related via the partial orders which they induce. For Lorentzian geometry we have causal strucure, for airplane boarding the natural partial order is blocking. We say that passenger X blocks passenger Y if it is impossible for passenger Y to reach his assigned row before passenger X (and others blocked by X) has sat down and cleared the aisle. Airplane boarding functions as a peeling process for the partial order defined by the blocking relation. At first passengers which are not blocked by any other passengers sit down, these passengers are the minimal elements in the blocking relation. In the second round passengers which are not blocked by passengers other than those of the first round are seated and so forth. Boarding time thus coincides with the size of the longest chain in the partial order. Let $X = (q,r)$ and $X' = (q + dq, r + dr)$, $dq > 0$, represent passengers with nearby coordinates. X blocks X' if $dr > 0$, however he may block X' even when $dr < 0$. Consider the time when passenger X arrives at his designated row. All passengers with row numbers beyond r which are behind passenger X in the queue but in front of passenger X' will occupy aisle space behind passenger X. The number of such passengers is roughly $N\alpha dq$. Each such passenger occupies u/w units

Discrete process	Events	Partial order	Number of rounds	Events in the same rounds
Space-time	random points	causal structure	length of longest geodesics	level curves
Disk scheduling	I/O requests	radial coordinate+unreacability in the same rotation	total service time	requests serviced in the same rotation
Polynuclear growth	nucleation events	layering	height of droplet	layers
Patience sorting	numbered cards in a deck	higher value later in the deck	longest increasing subsequence in the deck	cards in the same pile
Airplane boarding	passengers	passenger blocking	total boarding time	passengers who sit at the same time

TABLE I: A comparison of of different processes which are modeled by space-time

of aisle length where we take the basic aisle length unit to be the distance between rows. The row difference between X and X' is $-(N/b)dr$ thus passenger X is blocking passenger X', via the passengers which are behind him, roughly when $dq \geq -\alpha k dr$. Thus, we see that the metric corresponding to a boarding scenario is designed in such a way that the relation of blocking between passengers and the relation of causality between space-time points asymptotically coincide. The metric is then conformally scaled so that the volume form is proportional to the number of passengers. This procedure resembles the causal set theoretic approach to quantum gravity ,[1]. By the asymptotic equivalence between passenger blocking and causality we may replace in the analysis the airplane boarding process with the peeling process applied to the causal structure on points in space-time sampled with respect to the volume form. By a well known result on two dimensional Lorentzian metrics we can assume by a change of coordinates that the space-time is conformally flat on some domain (not necessarily the unit square). In light-like coordinates, chains in the causal relation in a conformally flat space-time coincides with increasing (upright) subsequences. The peeling process applied to the causal structure coincides with patience sorting which is a well known card game process which optimally computes the longest increasing subsequence in a permutation, [2], [3], [4]. Statement (A) then corresponds to a result on increasing subsequences of Deuschel and Zeitouni [5]. The global scaling constant $4D^2$ comes from a result of Vershik and Kerov [6]. Statement (B) follows from statement (A) and the fact that the level curves given by the equation $T(X) = \tau$ are Lorentz-orthogonal to maximal proper time trajectories. A previous computation of Aldous and Diaconis on pile sizes in patience sorting [3] corresponds to statement (B) when the space-time is assumed to be flat.

We may find chains of passenger blocking by considering a set of pointers in which each passenger points to the last passenger who blocked her way to the assigned row. The seating time of a passenger X will be D times the length of the trail of pointers starting from A. The construction of the metric ensures that this trail of pointers will cluster along a length maximizing geodesic trajectory ending at X.

So far we have assumed that D is fixed. It would seem more realistic to allow D to be a distribution. We can accommodate within our models delay distributions but the construction becomes somewhat less explicit. Given a distribution D, let W be the maximal weight of an upright sequence among N points uniformly chosen in the unit square with weights distributed according to D. By a result of Hammersley, [7] the random variable W will asymptotically concentrate on a value of the form $c_D\sqrt{n}$ for some constant c_D which depends on the distribution D. The metric which models airplane boarding when D is a distribution is given by $d\tau^2 = c_D^2 p(q,r)(dqdr + k\alpha(q,r)dq^2)$

Patience sorting which coincides with airplane boarding on conformally flat spaces is also known, [8], to be equivalent to the Polynuclear growth (PNG), [8], [9], which is a simple layer by layer surface growth model in dimension 1+1. PNG is in the Kardar-Parisi-Zhang (KPZ) universality class, [10]. In the mapping between PNG and airplane boarding, Nucleation events correspond to passengers. In previous studies of the PNG model (without sources) it was assumed that the probability of a nucleation event is uniform in space and time, which leads to flat space-time models on various domains D which depend on the initial conditions. When one assumes more generally an inhomogeneous environment in which nucleation rates may vary one obtains more general space-time models as in airplane boarding. The space-time model controls the macroscopic shape of the surface. In this setting the height of the surface at location x at time t will be given by $T(x,t)$, the length of the maximal time-like trajectory ending at (x,t) and the nucleation events whose spread is sufficient for creating the layers above position x are clustered along the length maximizing curve.

Using the relation between increasing subsequence and airplane boarding we can study the fluctuations in boarding time when the corresponding space-time is flat by reinterprating the corresponding results on increasing subsequences [11], [12]. A similar interpretation of results on increasing subsequences in the context of the PNG model is presented in [8]. We assume that the passengers are represented by N uniformly sampled points from a domain A in the unit square with metric $dqdr$. We assume that there is a unique maximal length trajectory in A, which we call C. When the endpoints of C are at corners on the boundary of the domain (for example if D is the unit square itself) the fluctuations are identical to those of the largest eigenvalue of a random N by N matrix in the Gaussian unitary ensemble (GUE). When only one endpoint is at a corner while the other lies on a straight line segment of the boundary, the fluctuations are identical to that of the largest eigenvalue of a matrix in the Gaussian symplectic ensemble (GSE). when the entire curve lies along the boundary of D between two corners we obtain the fluctuations of the largest eigenvalue in the gaussian orthogonal ensemble (GOE). All these cases appear naturally under various boarding process assumptions. In light of these results we expect the fluctuations of boarding time for the boarding process with general parameter settings (which do not correspond to flat metrics) to have features which are common to the various matrix models. In particular we expect

1) the order of magnitude for the fluctuations in boarding time about the value of $\sqrt{N}d(M)$ should be $N^{1/6}$.

2) The average value of the fluctuations should be negative.

3) The probability of a fluctuation of size $cN^{1/6}$ as $c \to \infty$ should have order of magnitude $e^{-\gamma c^{3/2}}$ for some constant γ.

Some initial experimental data using various boarding process parameters support these expected conclusions.

the fact that PNG is one of the processes in our class explains why the magnitude of fluctuations in airplane boarding are the same as those of KPZ in 1+1 dimensions as derived from the value of the dynamical scaling exponent $\beta = 1/3$.

boarding policies and the most natural instances of the disk scheduling problem lead to models with a continuous family of maximal length trajectories. For flat space-time models the fluctuations are on the order of $log(n)^{2/3} n^{1/6}$, since the fluctuations behave like an extreme value distribution on n^f independent variables Z_i where Z_i is the distribution describing the fluctuations associated with a single length maximizing trajectory and f is related to the transversal width of the fluctuations. We expect this to hold more generally for all models (not neccessarily flat) with families of maximizing curves.

Another problem which is known to be strongly related to the study of increasing sequences is the scheduling of I/O requests to a disk drive with a linear seek function [13], [14]. We can show that the problem of I/O scheduling can be modeled by a Lorentzian metric on the surface of the disk drive (or compact disk). A disk drive, or optical storage such as a compact disk, has the shape of an annulus $S^1 \times I$, where I is an interval. the annulus can be represented in normalized polar coordinates, say $0 \leq r \leq 1$, $0 \leq \theta \leq 1$ with the identification $(r, 0) = (r, 1)$. The problem of scheduling I/O requests can be presented as follows:

Given N I/O requests to different locations in the disk drive, find the optimal sequence in which the requests should be served so that total service time for all the requests is minimal.

Since not all data is equally popular we may assume that the locations of the N I/O requests are sampled with respect to a density $p(r, \theta) d\theta dr$ which quantifies the relative popularity of the data at location (r, θ). The metric on the annulus which models I/O scheduling is given by

$$d\tau^2 = p(r, \theta)(dr^2 - c d\theta^2)$$

Here c is a constant which depends on the radial speed of the read/write head of the disk drive and the rotational speed of the disk. In particular the optimal scheduling process (algorithm) of Andrews et al., [13], essentially coincides with the peeling process applied with respect to the causal structure of the space-time to the N request locations. In analogy with statement (A), the number of disk rotations which are required by the optimal scheduling algorithm to service all N requests is asymptotically $\sqrt{N}d(M)$. In the analogue of statement (B) dN refers to the number of I/O requests which are serviced between rotations $\sqrt{N}\tau$ and $\sqrt{N}(\tau + d\tau)$.

Disk scheduling can be mapped to PNG with a circular initial surface.

Statements (A) and (B) have several useful applications to the design and analysis of airplane boarding and disk scheduling strategies. A full description of the applications will be provided elsewhere. Here we want to present an application to the analysis of airline boarding strategies. Typically with the exception of first class passengers and passengers with special needs, airlines will attempt back to front boarding policies. Such policies initially allow passengers from the rear of the airplane to join the queue first and gradually allow more passengers (rows 40 and above, rows 30 and above) from the back towards the front. A boarding policy is given by a function $F(r)$ which represents the time in which passengers from row r are allowed to join the queue. Back to front policies correspond to non increasing functions $F(r)$. In order to specify a distribution $p(q, r)$, given a policy F, we also need a passenger reaction model which will tell us how passengers react to airline policies. One simple passenger reaction model assumes that passengers do not attempt to join the queue before they are allowed and join the queue at uniformly distributed times within a time units of being allowed to join. We may think of $1/a$ as a parameter which measures the attentivness of passengers. As a decreases the passengers become more attentive and the airline can exert more control via the policy F. This reaction model coupled with a policy F leads to a distribution $p(q, r)$ which is uniform on points (q, r) satisfying.

$$F(r) \leq q \leq F(r) + a$$

and vanishes outside this domain. We recall that the parameter k depends on interior design parameters of the airplane i.e., distance between rows and number of seats per row. When studying airline boarding strategies it is natural to fix k. Thus, given fixed values of a and k an airline policy specifies a model M_F and the problem of finding a good or optimal policy becomes the problem of minimizing $d(M_F)$ among all F.

When $k = 0$ we obtain flat space-time models which are supported on the domain $p > 0$ and the problem be-

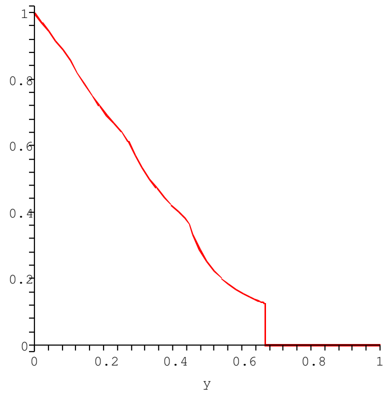

FIG. 1: Optimal policy for $k = 0$ and $a = 1/3$.

comes that of finding the domain with smallest diameter given a fixed a. The optimal policy for a fixed a is never unique. One can show that if F_1 and F_2 are optimal then so is $Max(F_1, F_2)$, leading to a unique maximal optimal policy F which is given by a sequence of hyperbola segments as pictured in figure 3 for the case $a = 1/3$.

The diameter of the corresponding model M_a is given by

$d(M_a) = \frac{2}{m}\sqrt{m+1-\frac{1}{a}}$

Where m is the unique integer satisfying

$\frac{2m+1}{(m+1)^2+m} \leq a < \frac{2m-1}{m^2+m-1}$.

In particular we conclude that as the attentivness of passengers increases the normalized boarding time as given by $d(M_a)$ tends to zero.

The same conclusion holds as long as $k < 1$, however, at $k = 1$ a phase transition occurs. To understand the phase transition we consider the family of linear back to front (attentivness dependent) boarding policies given by $F_a(r) = 1 - \frac{1}{(1-a)r}$. When $k < 1$ these policies satisfy $d(M_a) \longrightarrow 0$ as $T \longrightarrow 0$. On the other hand $d(M_T) \rightarrow \infty$ as $T \longrightarrow 0$ for $k > 1$. Geometrically the reason for the phase transition is that when $k > 1$, the antidiagonal $q + r = 1$ which was space-like for $k < 1$ becomes time-like and it's proper time increases as T tends to 0. The same phenomenon occurs for all back to front policies. Assuming the realistic values of 5-6 passengers per row, each occupying half the distance between successive rows we estimate that 2.5-3 is a realistic range for the parameter k. Based on this value and our analysis we conclude that simple back to front policies as practiced by many airlines are likely to be ineffective and even detrimental. These results are in line with results from simulation studies, [15], [16],[17]. In addition the results above indicate that local (in space and time) congestion which is reflected in the value of the parameter k, greatly affects boarding time. This suggests that airline policies should try to diffuse congestion by spacing passengers as much as possible at any given time. This can be achieved by applying multiple class policies which divide passengers into classes, say according to seat type. At any given time only one class is allowed to board. Multiple class policies effectively reduce the value of the local congestion parameter from k to k/l, where l is the number of classes. This explains the efficincy of multiple class policies as observed in simulation studies. our results provide a theoretical and quantitative basis for these conclusions.

In conclusion, we have shown that several processes which appear naturally in diverse contexts can be mapped to a peeling process with respect to the causal structure, acting on points in 2 dimensional space-time. The processes share common statistics which are strongly related to RMT. The analysis of these process carries applications to the design of good airplane boarding policies and I/O scheduling algorithms for disk drives and for the analysis of the macroscopic shape of polynuclear growth surfaces.

We are grateful to Perci Deift for many useful conversations and his constant encouragement.

[1] L. Bombelli, J. Lee, D. Meyer and R.D. Sorkin, Phys Rev. Lett. **59**, 521 (1987).

[2] C.L. Mallows, Bull. Inst. Math. Appl. **9**, 216 (1973).

[3] D. Aldous and P. Diaconis, Bull. A. Math. Soc. **36**, 413 (1999).

[4] M.L. Fredman, Disc. Math., **11**, 29 (1975).

[5] J.D. Deuschel and O. Zeitouni, Annals of Probability **23**, 852 (1987).

[6] A. Vershik and S. Kerov, Soviet Math. Dokl. **18**, 527 (1977).

[7] J.M. Hammersley, Proc. Sixth Berkeley Symp. Math. Statist. and prob. **1**, 345 (1972).

[8] M. Prahofer and H. Spohn Phys. Rev. Lett. **84**, 4882 (2000).

[9] P.Meakin *Fractals, Scaling and Growth Far From Equilibrium* (Cambridge University Press, Cambridge, England, 1998).

[10] M. Kardar, G. Parisi and Y.-C. Zhang, Phys. Rev. Lett. **56**, 889 (1986).

[11] J. Baik, P. Deift and K. Johansson, J. Am. Math. Soc. **12**, 1119 (1999).

[12] J. Baik and E. Rains, Duke J. of Math **109**, 205 (2001).

[13] M. Andrews, M. Bender and L. Zhang, Algorithmica **32**, 277 (2002).

[14] E. Bachmat, Proc. of Symp. Th. of Comp., 277 (2002).

[15] H. Van Landegham and A. Beuselinck, Euro. J. of Op. Res. **142**, 294 (2002).

[16] S. Marelli, G. Mattocks and R. Merry, Boeing Aero Magazine **1**, (2000).

[17] M. van den Briel, J. Villalobos and G. Hogg, Proc. of IERC (CD ROM only, 2003).