How many longest increasing subsequences are there?

Phil Krabbe,[1, *] Hendrik Schawe,[2, 1, †] and Alexander K. Hartmann[1, ‡]
[1]*Institut für Physik, Universität Oldenburg, 26111 Oldenburg, Germany*
[2]*Laboratoire de Physique Théorique et Modélisation,*
UMR-8089 CNRS, CY Cergy Paris Université, 95000 Cergy, France
(Dated: March 28, 2020)

We study the entropy S of longest increasing subsequences (LIS), i.e., the logarithm of the number of distinct LIS. We consider two ensembles of sequences, namely random permutations of integers and sequences drawn i.i.d. from a limited number of distinct integers. Using sophisticated algorithms, we are able to exactly count the number of LIS for each given sequence. Furthermore, we are not only measuring averages and variances for the considered ensembles of sequences, but we sample very large parts of the probability distribution $p(S)$ with very high precision. Especially, we are able to observe the tails of extremely rare events which occur with probabilities smaller than 10^{-600}. We show that the distribution of the entropy of the LIS is approximately Gaussian with deviations in the far tails, which might vanish in the limit of long sequences. Further we propose a large-deviation rate function which fits best to our observed data.

I. INTRODUCTION

Imagine a game of numbers: Given a sequence of n numbers, mark the largest subset of numbers such that every marked number is larger (or equal) to all marked numbers appearing left of it in the sequence. The marked numbers will be a *(weakly) increasing subsequence*. The number of marked elements is called the *length* l. If the subsequence maximizes l over all possible subsequences it is called a *longest (weakly) increasing subsequence* (LIS) [1]. An early study of this problem was by Stanisław Ulam [2] as a toy example to illustrate the Monte Carlo method in a textbook, which lead to its byname *Ulam's problem*. Though, it should be noted that in the same year Ref. [3] also discusses the connection of LIS to *Young tableaux*. Ulam's study found that LIS of random permutations have a mean length l which grows with the size of the sequence n as $\langle l \rangle = c\sqrt{n}$. The Monte Carlo simulations estimated $c \approx 1.7$ and in the years since then $c = 2$ was proven rigorously [4].

But the length of LIS of permutations attracted much more interest. In mathematics the whole distribution $p(l)$ was analyzed. First upper and lower tails were determined rigorously [5–7], and later it was proven that the central part is a Tracy-Widom distribution [8]. At the time this result was an unexpected connection between LIS and random matrix theory, where this Tracy-Widom distribution describes the fluctuations of the largest eigenvalues of the Gaussian unitary ensemble, i.e., an ensemble of Hermitian random matrices. In the following years it turned out that the LIS was an extremely simple model at the center of a growing class of seemingly unrelated problems. Beginning with a mapping of a 1+1-dimensional polynuclear growth model of the Kardar-Parisi-Zhang type onto LIS [9] a plethora of models were shown to exhibit the properties of LIS of random permutations, namely that their fluctuations are distributed according to one of the Tracy-Widom distributions. Examples range from other surface growth processes, like a direct mapping of a ballistic deposition model on LIS [10] or experimental observations of a Tracy-Widom distribution in the fluctuations of real surface growth [11], to the totally asymmetric exclusion process [12] and directed polymers [13]. For an overview and insight into the connections between these models, there are some review articles [14–16].

Recently, different ensembles of sequences than the random permutation were studied like random walks with different distributions of their jump lengths [17–20].

Besides its role in mathematics and physics, the LIS found applications in computer science, where it is suggested as a measure of *sortedness* of large amounts of data [21] or to find structures in time series while preserving privacy of the data, which is useful in the context of, e.g., fraud detection using financial data streams [22]. Also in bioinformatics the LIS found applications in the context of sequence alignment, e.g., for DNA or protein sequences [23].

Here, we are interested in another property of this famous problem. First, note that the LIS is not necessarily unique for any given sequence. For example, consider the sequence $\boldsymbol{\sigma} = (7, 9, 4, 1, 0, 6, 3, 8, 5, 2)$. While the length $l = 3$ of the longest increasing subsequence is uniquely defined, this sequence has $M = 7$ distinct LIS: $(4, 6, 8)$, $(1, 6, 8)$, $(0, 6, 8)$, $(1, 3, 8)$, $(0, 3, 8)$, $(1, 3, 5)$, $(0, 3, 5)$. As mentioned above, the length l is thoroughly studied, but about the number of distinct LIS M of a given sequence, only very little is known. Nevertheless, for example for the above mentioned application like determining sortedness or fraud detection, the actual number of distinct LIS will allow to estimate the reliability of decisions based on the LIS calculation much better. For this reason and to gain fundamental insight into the solution space structure, we study by computer simulations [24] here this quantity, namely its logarithm, i.e.,

---

[*] phil.krabbe@uol.de
[†] hendrik.schawe@cyu.fr
[‡] a.hartmann@uol.de

the entropy $S = \ln M$.

One of the few results is that the number of increasing subsequences of a fixed length grows exponentially in n [1, 25]. Although, this suggests that it is infeasible to count the LIS by enumeration, we will introduce in Sec. II B an algorithm to count the LIS efficiently without the need to enumerate them. Due to the exponential growth, it indeed makes sense to finally measure the entropy S. Since we want to explore the whole distribution of the entropy, including the tails of extremely rare events with probabilities of, say, 10^{-100}, we have to apply a sophisticated Markov chain sampling scheme, which will be explained in Sec. II C. Finally, Sec. III shows the results of our study before Sec. IV concludes this study. But first, we introduce the two ensembles we studied in Sec. II A.

II. MODELS AND METHODS

For completeness we define the LIS in a more formal way than in the introduction. Let $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \dots, \sigma_n)$ be a sequence of numbers. A LIS is a longest sequence $\lambda_{\boldsymbol{\sigma}} = (\sigma_{i_1}, \sigma_{i_2}, \dots, \sigma_{i_l})$ with $\sigma_{i_1} \le \sigma_{i_2} \le \dots \le \sigma_{i_l}$ such that $i_1 < i_2 < \dots < i_l \le n$. We denote by M the number of distinct sequences $\lambda_{\boldsymbol{\sigma}}$ fulfilling this property and $S = \ln M$ is the entropy.

A. Ensembles of random sequences

In this study, we scrutinize two ensembles of random sequences: First and more in depth, random permutations, for which an example with one LIS marked is visualized in Fig. 1(a).

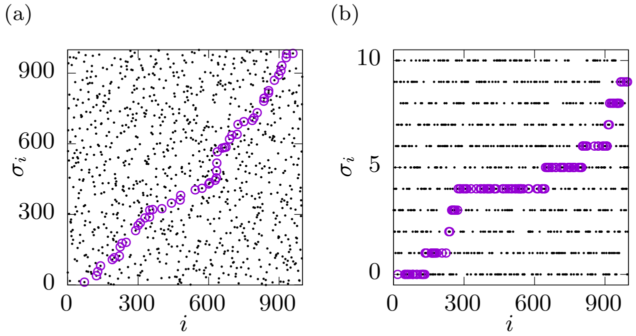

FIG. 1. (color online) Visualization of two sequences $\boldsymbol{\sigma}$. The horizontal axis shows the index i of the value σ_i. The elements belonging to one LIS are marked by circles. (a) Random permutation. (b) Random sequence with 11 distinct elements.

Second, we study a parameterized random sequence consisting of at most $K + 1$ distinct ordered elements. We call this "K ensemble". An example for $K = 10$ is shown in Fig. 1(b). In the limit $K = 0$, it consists only of identical elements and has therefore a unique LIS with a length of $l = n$. The other limit $K \to \infty$ consists of sequences with unique elements, which can be mapped to a permutation by replacing each element by its rank, which in turn will not change the LIS. Thus, we can interpolate with K between a non-degenerate LIS to the well known case of random permutations. Indeed, the length of the LIS of this ensemble was studied in Ref. [26].

As a technical remark, note that the algorithms explained in the following chapter find the strictly monotonic increasing subsequence, but for the K ensemble, we want to find the weakly increasing subsequence. With a simple mapping of the sequence (with elements from $\mathbb{N}_0$) to a new sequence $\alpha_i = \sigma_i + \frac{i}{n}$ of rational numbers we can apply algorithms for strict LIS on $\boldsymbol{\alpha}$ to find all weak LIS of $\boldsymbol{\sigma}$.

B. Counting the number of distinct LIS

Algorithms to find the length of the LIS are rather simple and there exists some variety. A popular choice is patience sort [27], which originally is a sorting algorithm especially suited for partially sorted data [28], but can be simplified to an efficient algorithm to find the length of the LIS of a given sequence in time $\mathcal{O}(n \ln n)$ [4]. But there are more alternatives, e.g., a fast algorithm in $\mathcal{O}(n \ln \ln n)$ [29, 30], approximate algorithms for sequences whose members can not be saved [31] or algorithms which are exact within a sliding window [32]. Even for the enumeration of LIS, there is literature introducing algorithms [29], which are able to, e.g., generate LIS with special properties [33].

Here, we introduce a method to count (and enumerate) distinct LIS of any sequence efficiently. Note that we do not claim to be the first to introduce an algorithm to count the number of LIS. Some of the existing enumeration algorithms could be extended with the same principle we use to allow for efficient counting. Also there is at least one algorithm description for counting the LIS in a well known programmer forum [34]. However, we could not find any reference to published literature. Therefore we show our approach, which is an extension of patience sort.

Like patience sort, this method takes elements sequentially from the front of the sequence and places them on top of a selected stack from a set $(s_1, \dots, s_k)$ of stacks, such that each stack s_i is sorted in decreasing order, i.e., the smallest element is on top of the stack, and the number of stacks is minimal. Thus, in the beginning there is just one stack containing the first element of the given sequence. This placement can be achieved by always placing the current element taken from the sequence on the leftmost stack, whose top element is larger than the current element. If this is not possible, i.e., if all top elements are smaller than the current element, one opens a new stack s_{k+1} right of the currently rightmost stack. Note that therefore the top elements of the stacks are ascendingly sorted and the correct stack of each element can be found via binary search. The number of stacks is equal to the length of the LIS l [4].

For counting the LIS, we need to extend this algorithm by introducing pointers. The basic idea is that for any LIS, exactly one number will be taken from each stack [35]. These pointers will take care of the order constraints in the following way: Each time an element is placed on a stack, pointers are added to some elements of the previous stack. This idea is already described in Ref. [4], but additionally to the pointers mentioned there, which will point to the currently topmost element of the previous stack, we also add pointers to all elements of the previous stack which are smaller than the current element. The meaning of such a pointer from any element σ_{j_1} to σ_{j_2} ($j_1 > j_2$) will be that in a LIS σ_{j_2} can appear before σ_{j_1}. An example structure is shown in Fig. 2.

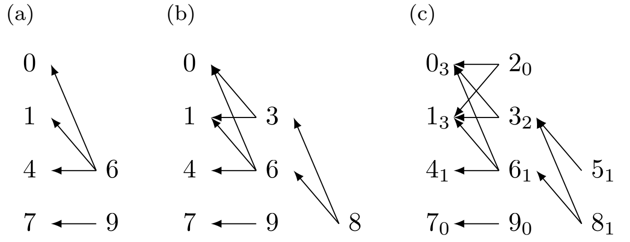

FIG. 2. Construction of the DAG for the sequence $(7, 9, 4, 1, 0, 6, 3, 8, 5, 2)$. Stacks grow upwards. (a) Partial DAG for $(7, 9, 4, 1, 0, 6)$. (b) Partial DAG for $(7, 9, 4, 1, 0, 6, 3, 8)$. (c) Complete DAG with annotations (subscript) labeling the number of paths to reach the corresponding element from the rightmost stack. Summing the subscripts of the leftmost stack yields the total number of paths originating from the rightmost stack, i.e., the number of all LIS, here $M = 7$.

The set of all pointers, i.e., edges, forms a directed acyclic graph (DAG). The DAG can be used to enumerate all LIS by following all paths originating from any element of the rightmost stack. This will yield all LIS in reversed order. For our purpose, we just have to count all paths originating from the rightmost stack. Therefore we propagate the information by how many paths an element can be reached through the DAG. All elements of the rightmost stack are initialized with 1. The elements of the stack left are assigned the sum of all incoming edges. This is repeated until the leftmost stack is reached. The sum of all paths ending in elements of the leftmost stack is the total number of LIS.

To estimate the run time, note that we have to iterate over all incoming edges, of which there are at most $\mathcal{O}(n^2)$ in a DAG with n nodes. Also the construction takes the maximum of the number of edges for constructing the pointers and $\mathcal{O}(n \ln n)$ for constructing the stacks, such that the runtime of this algorithm is $\mathcal{O}(n^2)$ in the worst case. Note however that typical DAG generated here have far fewer edges. We observed that typically, the length of a LIS and therefore the number of stacks is $\mathcal{O}(\sqrt{n})$. Each stack can only be connected to the previous stack. Assuming that stacks are typically of size $\mathcal{O}(\sqrt{n})$, this leads to at most $\mathcal{O}(n)$ edges between each pair of neighboring stacks and therefore $\mathcal{O}(n\sqrt{n})$ total edges.

C. Sampling rare events

Using the algorithm above, we can determine the number of LIS M for arbitrary sequences. Therefore, generating random sequences allows us to sample $S = \ln M$, build histograms from the samples and estimate the distribution $p(S)$ from them. But to observe any event which occurs with a probability of r, we would have to generate $\mathcal{O}(1/r)$ samples and $\mathcal{O}(1/r^2)$ to reduce the statistical error enough to determine the probability with reasonable accuracy. Since we would like to know the probability distribution also in the extremely rare event tails, we have to use a more sophisticated method than this proposed *simple sampling*.

Our approach is to bias the ensemble in a controlled way towards extremely improbable configurations, gather enough samples there and correct the bias afterwards. This will lead to small statistical errors across large parts of the support. This method [36] was successfully applied in a wide range of problems from graph theory [37, 38], over stochastic geometry [39], nonequilibrium work distributions [40], the Kardar-Parisi-Zhang equation [41] to the exploration of the tails of the distribution of the LIS's length for random permutations and random walks [19].

The exact method is inspired by equilibrium thermodynamics, where the Metropolis algorithm [42] is used to generate samples of systems in the canonical ensemble at some *temperature* T, which governs the typical values of the *energies* observed in this system. Here, we identify the energy with our observable of interest S. This allows us to use the "temperature" parameter to bias the generated states towards improbable values of S.

This method builds a Markov chain consisting of sequences $\boldsymbol{\sigma}^{(i)}$, where i is the step counter of the chain. For each step in the chain, from the present sequence $\boldsymbol{\sigma}^{(i)}$ a trial sequence $\boldsymbol{\sigma}'$ is constructed by performing some changes to $\boldsymbol{\sigma}^{(i)}$. For the standard ensemble of random permutations, where we have performed large-deviation simulations, we used as change to swap two elements. The trial sequence is accepted, i.e., $\boldsymbol{\sigma}^{(i+1)} = \boldsymbol{\sigma}'$ with the Metropolis acceptance probability $P_{\text{acc}} = \min\left(1, e^{-\Delta S/T}\right)$ depending on the temperature T and the change ΔS between $\boldsymbol{\sigma}'$ and $\boldsymbol{\sigma}^{(i)}$. Otherwise the previous sequence is repeated in the chain, i.e., $\boldsymbol{\sigma}^{(i+1)} = \boldsymbol{\sigma}^{(i)}$ This procedure is sketched in Fig. 3 and will eventually result in sequences $\boldsymbol{\sigma}$ occurring in the chain which are distributed according to

$$Q_T(\boldsymbol{\sigma}) = \frac{1}{Z_T} e^{-S(\boldsymbol{\sigma})/T} Q(\boldsymbol{\sigma}) \tag{1}$$

where $Q(\boldsymbol{\sigma})$ is the natural distribution of sequences, which we would obtain from simple sampling and Z_T

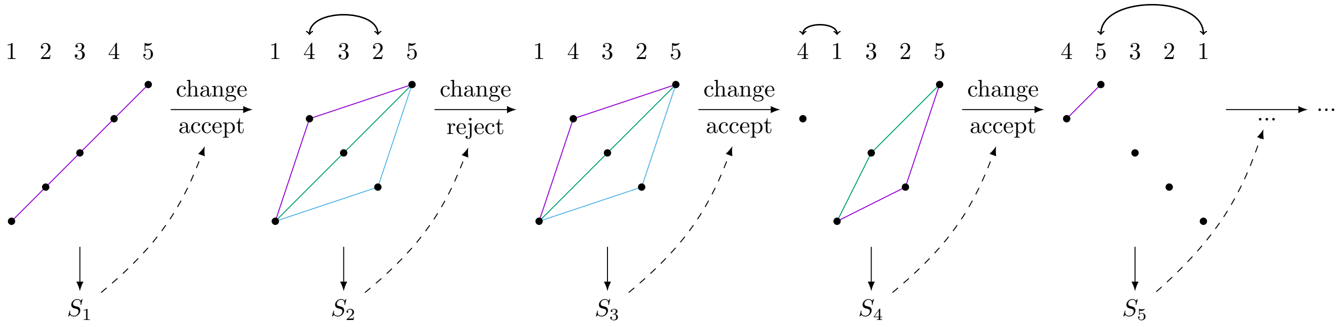

FIG. 3. (color online) Sketch of a Markov chain of LIS realizations generated by swaps of two random elements of the permutation. All distinct LIS are marked by lines of distinct color (shade). The acceptance of a sequence as the next sequence of the Markov chain is dependent on the number of LIS in the realization M, since the energy is identified with $S = \ln M$.

is the partition function of our artificial temperature ensemble. From here it is just a question of elemental algebra to connect our estimates of the probability density function in the artificial temperature ensemble $p_T(S)$ to the distribution of the unbiased ensemble we want to study $p(S)$:

$$\begin{aligned} p_T(S) &= \sum_{\{\boldsymbol{\sigma}|S(\boldsymbol{\sigma})=S\}} Q_T(\boldsymbol{\sigma}) && (2)\\ &= \sum_{\{\boldsymbol{\sigma}|S(\boldsymbol{\sigma})=S\}} \frac{1}{Z_T} e^{-S(\boldsymbol{\sigma})/T} Q(\boldsymbol{\sigma}) && (3)\\ &= \frac{1}{Z_T} e^{-S(\boldsymbol{\sigma})/T} p(S). && (4)\end{aligned}$$

Depending on the value of T, a simulation will generate data for S in a specific intervall. Thus, to obtain the distribution $p(S)$ over a large range of the support, we performed simulations for many values of T. This requires finely tuned values of the temperatures. The ratio of all constants Z_T can be obtained from overlaps of p_{T_i} and p_{T_j}, since the actual distribution needs to be unique, i.e.,

$$p_{T_j}(S) e^{S/T_j} Z_{T_j} = p_{T_i}(S) e^{S/T_i} Z_{T_i}. \qquad (5)$$

We used in the order of 30 temperatures per size n, where larger sizes typically required more temperatures. Also, like for all Markov chain Monte Carlo one has to carefully ensure the equilibration of the process and discard sequences of the chain which are still correlated too much with previous sequences. Note that equilibration can be ensured rather conveniently [36], by performing two sets of simulations starting with very different initial sequences, with low and with high values of S, respectively. The Markov chains can be considered being equilibrated, when the values of S agree within fluctuations between the two sets.

III. RESULTS

First, we study the behavior of typical sequences for the two ensembles. In the second part, we will investigate the large-deviation behavior of the standard permutation ensemble.

A. Typical behavior

To investigate the typical behavior of the permutation ensemble, we consider different system sizes up to large sequences of $n = 524288 = 2^{19}$ elements. The estimated probability density functions $p(S)$ of the LIS-entropy of permutations is shown in the range of typical probabilities in Fig. 4(a). These data are collected over 10^6 samples for each system size. Clearly, the mean value and width of the distribution increase with n.

Indeed, we observe for the mean a growth of the form

$$\langle S \rangle = c\sqrt{n}. \qquad (6)$$

(Also see below in Fig. 5(b).) Note that the fits resulted in rather large reduced χ^2 goodness of fit values (caused by the very high precision of the measured means) suggesting that there are corrections to this form for finite sizes. Our best estimate for the prefactor under the assumption that the above relation is correct, is $c \approx 0.347$. Also, for the standard deviation we observe a similar simple relation of $\sigma_S = b\sqrt[4]{n}$ with $b \approx 0.49$.

We notice that the growth of the mean entropy $\langle S \rangle \approx c\sqrt{n}$ and of the mean count $\langle M \rangle \approx e^{c_2\sqrt{n}}$ with $c_2 \approx 0.44$ (not shown) estimated from our data, follow the same behavior as the mean number of increasing sequences (IS) of length $2\sqrt{n} = \langle l \rangle$ [25, Eq. (11.5)], $\langle m \rangle \approx \left(\frac{e}{2}\right)^{4\sqrt{n}} = e^{4\ln(e/2)\sqrt{n}}$. Thus, since $c_2 < 4\ln(e/2) \approx 1.2$, our numerical results suggest the actual mean count of LIS to be much lower (and $e^{\langle S \rangle}$ even lower in accordance to Jensen's inequality). In other words, the number of *longest* increasing subsequences is exponential lower than

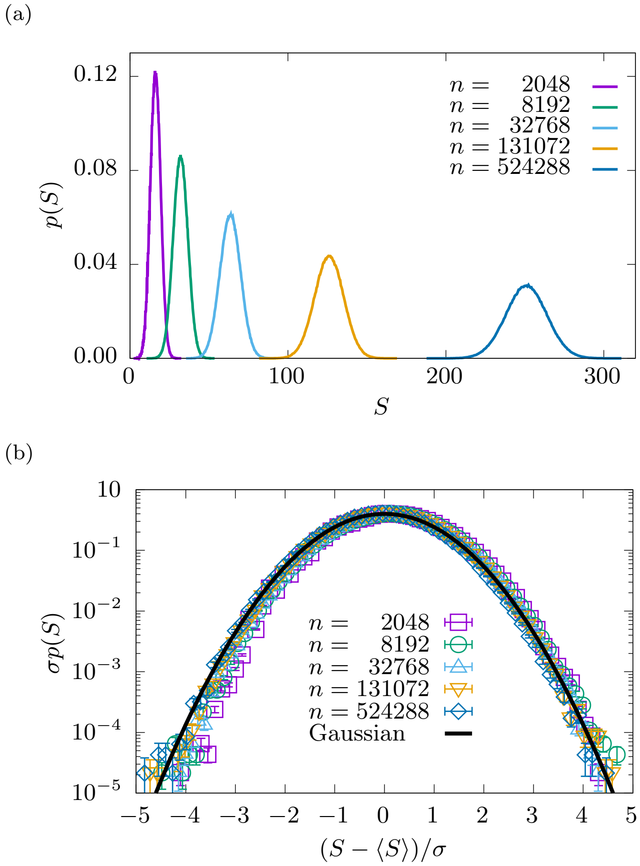

FIG. 4. (a) (color online) Probability density $p(S)$ of the entropy S for random permutations of different length n, obtained with simple sampling. (b) Probability densities $p(S)$ collapse on an approximate standard Gaussian shape for multiple system sizes if shifted by their mean $\langle S \rangle \approx 0.347\sqrt{n}$ and scaled with their width $\sigma \approx 0.49\sqrt[4]{n}$. Note the logarithmic vertical axis.

number of increasing subsequences of the same length. This can be understood in the following way: We consider IS of given length $\bar{l}$, which is the average LIS length for this value of n. Now, looking at the ensemble of sequences, some will have a LIS length $l < \bar{l}$. For them, there are no IS of length $\bar{l}$, so they will not contribute any IS to the average. This will be a fraction of sequences. Some sequences will have LIS length of $l = \bar{l}$, they will contribute all their LISs to the average. Finally, some fraction of sequences will have a LIS length $l > \bar{l}$. Here, all subsequences of length $\bar{l}$ of all LIS will be IS contributing to the average. Since there are exponentially many subsequences (and maybe even more IS which are not subsequences of a LIS), they will dominate the average number of IS, thus leading to a stronger exponential growth as compared to the average number of LIS.

We use our estimates for the mean and standard deviation to rescale the distributions for different system sizes in Fig. 4(b) and observe a collapse on a shape which can be approximated well by a standard Gaussian. Especially, the strongest deviations from this scaling form occur for small sizes, while the larger sizes seem to converge to the limiting shape. This is expected since the corrections to the scaling we used, which are mentioned above, should be stronger for smaller values of n. We backed this observation by classical normality tests [43–45], which are able to distinguish this distribution from a normal distribution with very high confidence at small system sizes, but become less confident for the largest system sizes (details not shown here). Especially the weak Kolmogorov-Smirnov test is not able to distinguish the distributions from a normal distribution with a significance level below 10% for all sizes $n \geq 65536$ of for our sample of 10^6 realizations.

Due to our limited sample size, the tails of the measured distribution are subject to statistical errors. In Sec. III B we will present higher quality data for the far tails to show that the approximation by a Gaussian shape is valid deep into the tails. Even for extreme rare events we can not exclude the possibility that the distribution converges to a Gaussian in the large-n limit.

Next, we look at the ensemble of random sequences with a limited number of $K + 1$ distinct elements. For constant values of K Fig. 5(a) shows the average entropy. The trivial case of $K = 0$, which allows only one LIS of length $l = n$, corresponds to an entropy of $S = 0$ and is not visualized. The case $K = 1$, which consists of sequences containing two distinct elements, has a low and interestingly almost n-independent entropy. There are typically only one or two distinct LIS in such sequences independent of the length of the sequence. Our data for larger values of K show two phenomena. First, larger values of K lead to larger entropies and second, the dependency of the entropy on the length of the sequence diminishes for the limit of large n, i.e., for each fixed K there should be a limit entropy approached for $n \to \infty$. Indeed, fits of a function $f(n) = b + \frac{a}{\ln(n-c)}$ with the limiting value $\lim_{n\to\infty} f(n) = b$ to our data confirm this guess for all values of K we considered. Note that the shape of the fitting form is purely heuristical. We first tried standard shapes like approaching a constant with a power law or an exponential, but they did not work out well.

Since b seems to grow roughly linear with K (not shown), this leads to the conjecture that for $K \propto n$ the saturation of the entropy should not occur and instead grow with the size n. Especially, we observed this behavior for the permutation case, which is identical to the $K \to \infty$ limit, as explained in Sec. II A. Indeed, in Fig. 5(b) we can observe a quick convergence with increasing n to the behavior of the random permutation. Our conjecture Eq. (6) for its growth is visualized as a line.

B. The Far Tails

In this section we study the distribution of the entropy of the LIS for the random permutation ensemble with a

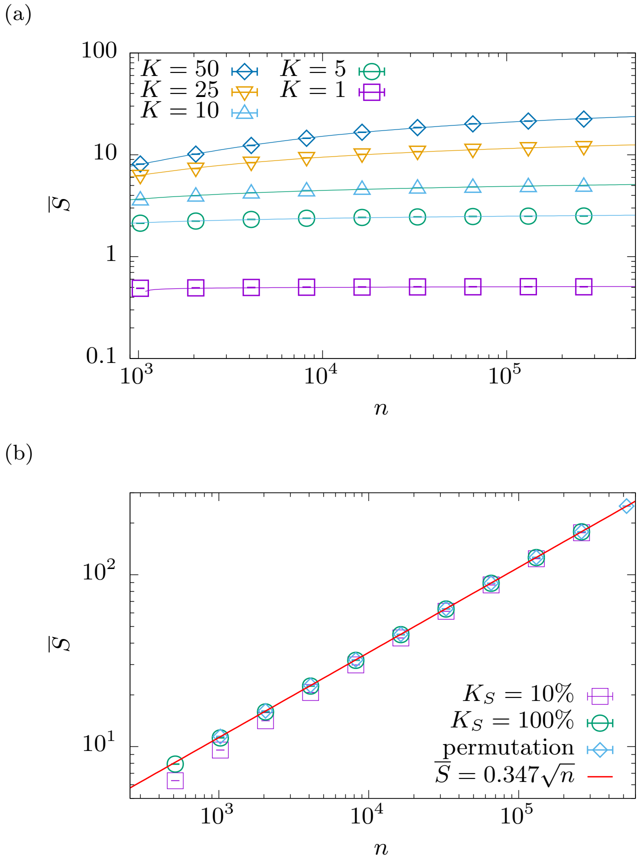

FIG. 5. (color online) Average entropy $\overline{S}$ as a function of the sequence length n, for the permutation ensemble and for the K ensemble with different values of K. Error bars are usually smaller than the width of the lines. (a) K ensemble with constant K. Lines are fits of the form $f(n) = b + \frac{a}{\ln(n-c)}$ (b) Permutation ensemble and with $K \propto n$. The line is the growth observed for random permutations $\langle S \rangle \approx 0.347\sqrt{n}$.

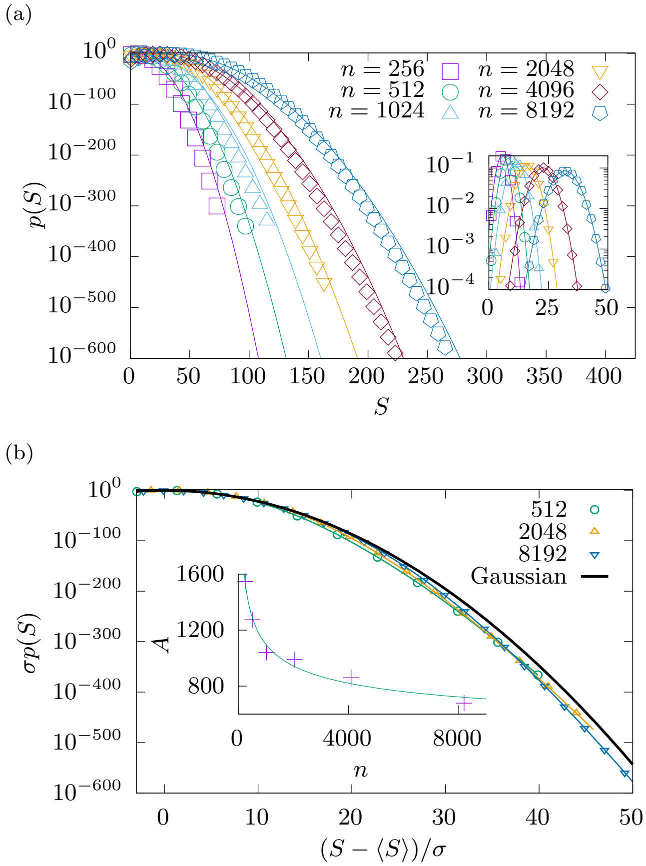

FIG. 6. (color online) (a) Probability density $p(S)$ for multiple system sizes with extremely high precision data for the far tails. The inset shows a zoom on the high probability region. The lines are fits to Gaussian distributions, which fit very well in the high-probability region, but do not describe the whole tails of the distribution. (b) Same rescaling of the axes as Fig. 3(b). This shows that the different system sizes move towards the Gaussian for larger sizes. The lines are linear interpolations of all available data points, not all of them are shown as symbols for clarity. The inset shows the area between the logarithm of the rescaled distribution and the logarithm of a standard normal distribution with a fit used for extrapolation.

focus on the far tails. Due to the much larger numerical effort, we are able to show results up to sequence lengths of $n = 8192$.

The data presented in the previous section, which was obtained via simple sampling, resulted in a distribution which appeared to be very well approximated by a Gaussian. We want to investigate whether this is still true when including our high-precision estimates of the far tails. Here, we can observe a slightly faster than Gaussian decay, see Fig. 6(a). There the distribution for a selection of sequence lengths n is shown including the far tail and fitted with (normalized) Gaussians. While they describe the high-probability region of the distribution (shown in the inset) very well, the deviation becomes stronger for increasingly rare events.

To test whether this deviation remains in the $n \to \infty$ limit, we rescale the distributions, like in Fig. 4(b), to be independent of system size, see Fig. 6(b). First, we see that this collapse does not work as well in the far tails, as it does in the high-probability region. Interestingly, there is a crossing for different system sizes. Left of the crossing larger sizes tend towards the Gaussian, which hints that for larger sizes the Gaussian approximation becomes better even in the intermediate tails. Careful examination of the crossing shows that its position is dependent on the system size and larger systems cross farther on the right than smaller systems. This is again a hint that the Gaussian approximation becomes valid over larger ranges of the distribution for larger system sizes.

To quantify this observation, we can not use classical statistical tests like we could do it for the data we obtained via simple sampling. Instead we will use a crude estimate of similarity, similar to one already used in [46]. We compare the area A between the logarithm of the scaled empirical distributions $p_s(x) = \sigma p(S)$, where $x = (S - \langle S\rangle)/\sigma$ (cf. Fig. 6(b)), and the logarithm of a standard normal probability density function p_{G} to estimate whether they become more similar for larger sizes. To be comparable across all system sizes, we limit this difference to the largest range of the horizontal axis, for which we have data for all sizes, i.e.,

$$A = \int_0^{40} |\ln p_s(x) - \ln p_{\mathrm{G}}(x)| \, \mathrm{d}x. \qquad (7)$$

Using this method we observe a strictly decreasing area, as shown in the inset of Fig. 6(b). If we extrapolate it using a power law with offset $A(n) = c + an^{-b}$ we obtain a result for the offset $c = 344 \pm 360$, which is within errorbars consistent with an offset of 0, i.e., consistent with a convergence to a Gaussian. However, since the constants of our scaling assumption are tainted with hard to quantify errors, this should be interpreted as a trend and not as a rigorous result. Nevertheless this result means that we can not exclude the possibility that the distribution will be Gaussian in the right tail in the limit of $n \to \infty$.

Next, we can use our empirical data of the distribution to test whether a *large deviation principle* holds, that is whether the behavior of the distribution can be expressed by a *rate function* Φ in the $n \to \infty$ limit, defined by $\Phi(s) = -\lim_{n\to\infty} \frac{1}{n} \ln p_n(sS_{\max,n})$ [47], where $S_{\max,n}$ is the maximum possible value for a given value of n and therefore $s \in [0, 1]$. This rescaling with the maximum value, i.e., $s = S/S_{\max,n}$, is done to describe the largest fluctuations by one size-independent function $\Phi(s)$. Since we have the distributions $p_n(S)$ for multiple finite n, we can calculate *empirical rate functions* $\Phi_n(s)$ for each n and extrapolate whether they converge to a limiting curve, which is a strong hint that this curve is the size-independent rate function $\Phi(s)$, which would establish a large deviation principle. If a rate function exists, it governs the fluctuations around the mean. For example, the existence of a rate function with mild properties implies the law of large numbers and the central limit theorem for the corresponding process. For brevity, we will omit n-subscripts for $S_{\mathrm{max,n}}$ and $p_n(S)$.

To analyze our data, we have to determine $S_{\max}$ for the LIS. A maximum entropy is achieved if for many groups of elements, one can choose independently between different elements. Thus, consider a sequence, which consists of groups of k decreasing elements, followed by k decreasing elements, which are larger than all elements before and so on. An example for $n = 9$ and $k = 3$ is $(3, 2, 1, 6, 5, 4, 9, 8, 7)$. In this case a LIS would have length n/k and can contain for each block of k an arbitrary element, resulting in $M = k^{n/k}$ distinct LIS. The entropy $S = \ln M = \frac{n}{k} \ln k$ is maximized at $k = e$, and since we are limited to integer k, at $k = 3$. This results in a maximum entropy of $S_{\max} = n \ln 3/3$, i.e., linear in n.

In Fig. 7(a) the empirical rate functions are visualized, but no convergence to a common tail is visible. The best common tail we can generate, happens for a slightly modified rate function with an unusual exponent $\Phi_u(s) = -\lim_{n\to\infty} \frac{1}{n^{3/2}} \ln p(S)$. Note however, that such an exponent is not out of the question. For example, the rate function of the right tail of the distribution for the rescaled length $\tilde{l} = l/\sqrt{n}$ of LIS behaves as $\Phi(\tilde{l}) = -\lim_{n\to\infty} \frac{1}{n^{1/2}} \ln p(l)$ [7]. Our result is shown in Fig. 7(b) in double logarithmic scale to emphasize the collapse in the right tail on a power law $\sim s^{\kappa}$ with a slope of $\kappa \approx 2$, consistent with the before observed almost Gaussian right tail.

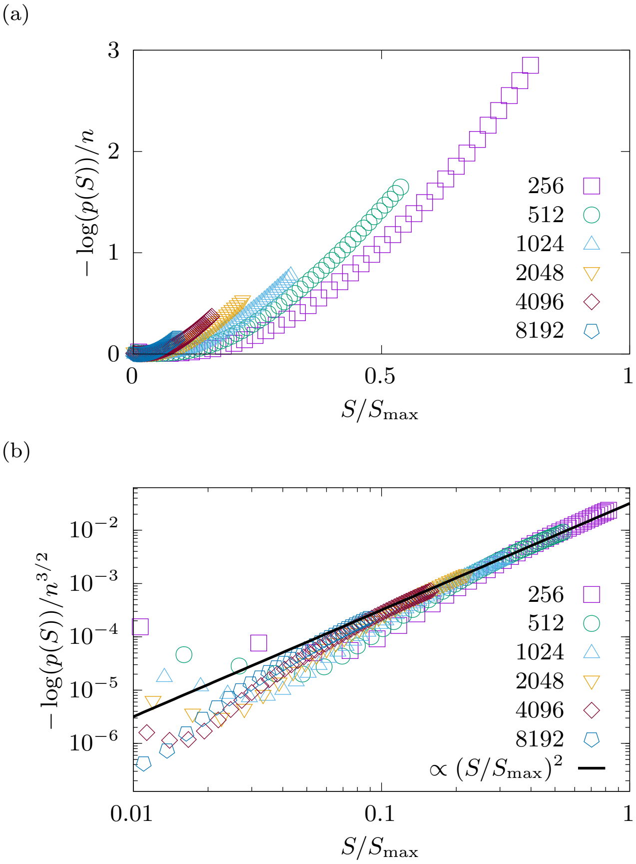

FIG. 7. (color online) (a) Usual empirical rate function $\Phi(s) = -\lim_{n\to\infty} \frac{1}{n} \ln p(S)$. No convergence is visible, the curves shift to the left with increasing sequence length n. (b) Empirical rate function with unusual exponent $\Phi_u(s) = -\lim_{n\to\infty} \frac{1}{n^{3/2}} \ln p(S)$ for the random permutation case in log-log scale to emphasize the convergence to a common tail with a power-law shape.

Finally, we want to understand what leads to sequences with atypically many or few distinct LIS. For this purpose we used the sequences generated by simple sampling and by the large-deviation approach to study their correlation with the length of the corresponding LIS. This might give insight into qualitative mechanisms governing the degeneracy of the LIS. This correlation is visualized in Fig. 8 for random permutations of length $n = 8192$. Typical permutations have a LIS-entropy around $\langle S\rangle \approx 35$

with a typical length of $\langle l\rangle \approx 180$, marked by darker gray points. Apparently, the degeneracy of the LIS is uncorrelated with its length for low and intermediate values of S. Though, extremely degenerate LIS necessitate longer LIS. This is somewhat counter intuitive, since there are more increasing subsequences of shorter length [1]. We assume therefore that the mechanism here is that these rare sequences have a structure which is in some sense modular (cf. Sec. III B for the configuration of maximum entropy) to allow for many almost identical LIS. These may differ independently in many places and therefore can combinatorically combine such small differences. Then, since a longer LIS has more members, the combinatorial character leads to an entropy advantage for long LIS. This higher number of combinatorial possibilities becomes necessary at some point to support even more degenerate LIS, thus we see a very strong correlation for extremely high values of S. However, we assume that for very long LIS, the entropy has to decrease again. In the extreme case of $l = n$ the sequence has to be sorted and can only contain a single LIS. Since we do not observe very long LIS in our sampled data, it means that they are combinatorially suppressed. In order to have access to this region, one would have to perform a biased sampling with respect to the length and measure the entropy, or, even better, measure the two-dimensional distribution $p(S, l)$ by a two-temperature large-deviation approach, which would be numerically very demanding.

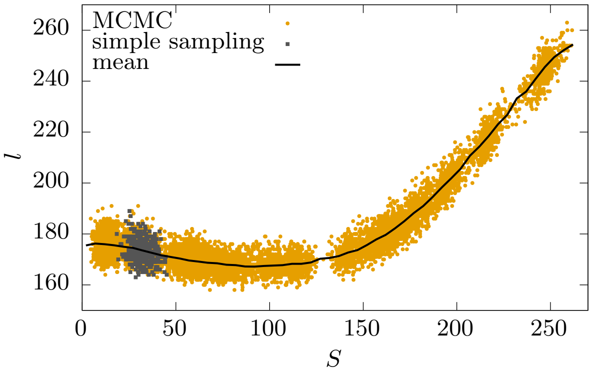

FIG. 8. (color online) The length of the LIS l as function of the entropy S for a sequence length $n = 8192$. The dark gray data points are gathered using simple sampling and represent typical sequences. The other data points are collected with the Markov chain Monte Carlo (MCMC) method described in Sec. II C and represent extremely rare sequences with atypical entropy.

IV. CONCLUSIONS

Here, we studied the entropy S of longest increasing subsequences of random permutations by counting the number of distinct LIS. Using an extension of the patience-sort algorithm, this can be readily obtained for any given sequence. Especially we applied Markov chain Monte Carlo techniques to explore the far tail of the probability distribution of S in the regime of extremely rare events with probabilities less than 10^{-600}.

Concerning the typical behavior, we found that the average entropy grows as a square root in the length of the permutation, i.e., the number of LIS grows exponentially, as expected. The fluctuations of the entropy are in good approximation Gaussian, but show deviations from this shape in the far tails.

Further, we use the data of the far tails to empirically scrutinize the rate function, the central piece of large-deviation theory. For the right tails we propose a rate function with an unusual exponent $\Phi\left(S/S_{\max}\right) = -\lim_{n\to\infty} \ln p\left(S\right)/n^{3/2} \sim \left(S/S_{\max}\right)^2$, towards which the right tails of all studied system sizes seem to converge. This means, the standard large-deviation principle, where one would see a convergence with a factor $1/n$ instead of $1/n^{3/2}$, does not hold, but still the tails of the distribution can be described by some rate function. Note that for the distribution of LIS lengths also a rate function with a factor different from $1/n$ was found in previous work.

Additionally to random permutations, we studied an ensemble with a limited amount of distinct elements in the sequences. For fixed number of distinct elements, we observed that S converges to a constant value which is independent of the sequence length for long sequences. For any number of distinct symbols, which is proportional to the sequence length, this will converge to the same LIS-entropy as random permutations for large system sizes.

Also, the datastructure used to count the LIS can be used to perform unbiased sampling of all LIS, which is a line of research the authors are working on right now. Here, also other ensembles of sequences could be of interest, like one-dimensional random walks. Finally, for future research it could be interesting, yet numerically extremely demanding, to study the two-dimensional distributions like $p(S, l)$.

ACKNOWLEDGMENTS

HS would like to thank Naftali Smith for interesting discussions about LIS. The simulations were performed at the HPC Cluster CARL, located at the University of Oldenburg (Germany) and funded by the DFG through its Major Research Instrumentation Programme (INST 184/108-1 FUGG) and the Ministry of Science and Culture (MWK) of the Lower Saxony State.

---

[1] D. Romik, *The Surprising Mathematics of Longest Increasing Subsequences* (Cambridge University Press, USA, 2015).

[2] S. M. Ulam, in *Modern Mathematics for the Engineer: Second Series*, Dover Books on Engineering Series, edited by E. Beckenbach and M. Hestenes (Dover Publications, Incorporated, 2013) Chap. 11, pp. 261–281.

[3] C. Schensted, Canadian Journal of Mathematics **13**, 179–191 (1961).

[4] D. Aldous and P. Diaconis, Bulletin of the American Mathematical Society **36**, 413 (1999).

[5] T. Seppäläinen, Probability Theory and Related Fields **112**, 221 (1998).

[6] B. F. Logan and L. A. Shepp, Advances in Mathmatics **26**, 206 (1977).

[7] J.-D. Deuschel and O. Zeitouni, Combinatorics, Probability and Computing **8**, 247 (1999).

[8] J. Baik, P. Deift, and K. Johansson, Journal of the American Mathematical Society **12**, 1119 (1999).

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

[10] S. N. Majumdar and S. Nechaev, Physical Review E **69**, 011103 (2004).

[11] K. A. Takeuchi and M. Sano, Phys. Rev. Lett. **104**, 230601 (2010).

[12] K. Johansson, Communications in Mathematical Physics **209**, 437 (2000).

[13] J. Baik and E. M. Rains, Journal of Statistical Physics **100**, 523 (2000).

[14] T. Kriecherbauer and J. Krug, Journal of Physics A: Mathematical and Theoretical **43**, 403001 (2010).

[15] S. N. Majumdar, in *Complex Systems: Lecture Notes of the Les Houches Summer School 2006*, Les Houches, edited by J. Bouchaud, M. Mézard, and J. Dalibard (Elsevier Science, 2006) Chap. 4.

[16] I. Corvin, Random Matrices: Theory and Applications **01**, 1130001 (2012).

[17] O. Angel, R. Balka, and Y. Peres, Mathematical Proceedings of the Cambridge Philosophical Society **163**, 173 (2017).

[18] J. R. G. Mendonça, Journal of Physics A: Mathematical and Theoretical **50**, 08LT02 (2017).

[19] J. Börjes, H. Schawe, and A. K. Hartmann, Phys. Rev. E **99**, 042104 (2019).

[20] J. R. G. Mendonça, H. Schawe, and A. K. Hartmann, Phys. Rev. E **101**, 032102 (2020).

[21] P. Gopalan, T. S. Jayram, R. Krauthgamer, and R. Kumar, in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '07 (Society for Industrial and Applied Mathematics, USA, 2007) p. 318–327.

[22] L. Bonomi and L. Xiong, Transactions on Data Privacy **9**, 73 (2016).

[23] H. Zhang, Bioinformatics **19**, 1391 (2003).

[24] A. K. Hartmann, *Big Practical Guide to Computer Simulations* (World Scientific, Singapore, 2015).

[25] J. M. Hammersley, in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Theory of Statistics* (University of California Press, Berkeley, Calif., 1972) pp. 345–394.

[26] C. Houdré and T. J. Litherland, "On the longest increasing subsequence for finite and countable alphabets," in *High Dimensional Probability V: The Luminy Volume*, Collections, Vol. Volume 5, edited by C. Houdré, V. Koltchinskii, D. M. Mason, and M. Peligrad (Institute of Mathematical Statistics, Beachwood, Ohio, USA, 2009) pp. 185–212.

[27] C. L. Mallows, SIAM Review **5**, 375 (1963).

[28] B. Chandramouli and J. Goldstein, in *ACM SIGMOD International Conference on Management of Data (SIGMOD 2014)* (ACM SIGMOD, 2014).

[29] S. Bespamyatnikh and M. Segal, Information Processing Letters **76**, 7 (2000).

[30] M. Crochemore and E. Porat, Information and Computation **208**, 1054 (2010).

[31] A. Arlotto, V. V. Nguyen, and J. M. Steele, Stochastic Processes and their Applications **125**, 3596 (2015).

[32] M. H. Albert, A. Golynski, A. M. Hamel, A. López-Ortiz, S. Rao, and M. A. Safari, Theoretical Computer Science **321**, 405 (2004).

[33] Y. Li, L. Zou, H. Zhang, and D. Zhao, Proc. VLDB Endow. **10**, 181–192 (2016).

[34] A. Salauyou, `https://stackoverflow.com/questions/22923646/number-of-all-longest-increasing-subsequences/22945390#22945390`.

[35] Other numbers is the same stack appear earlier in the same sequence but are larger, or they appear later but are smaller.

[36] A. K. Hartmann, Phys. Rev. E **65**, 056102 (2002).

[37] A. K. Hartmann, The European Physical Journal B **84**, 627 (2011).

[38] H. Schawe and A. K. Hartmann, The European Physical Journal B **92**, 73 (2019).

[39] H. Schawe, A. K. Hartmann, and S. N. Majumdar, Phys. Rev. E **97**, 062159 (2018).

[40] A. K. Hartmann, Phys. Rev. E **89**, 052103 (2014).

[41] A. K. Hartmann, P. L. Doussal, S. N. Majumdar, A. Rosso, and G. Schehr, Europhys. Lett. **121**, 67004 (2018).

[42] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, The journal of chemical physics **21**, 1087 (1953).

[43] R. D'Agostino and E. S. Pearson, Biometrika **60**, 613 (1973).

[44] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes 3rd edition: The art of scientific computing* (Cambridge university press, 2007).

[45] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. Jarrod Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. Carey, İ. Polat, Y. Feng, E. W. Moore, J. Vand erPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and S. . . Contributors, arXiv e-prints , arXiv:1907.10121 (2019), arXiv:1907.10121 [cs.MS].

[46] H. Schawe and A. K. Hartmann, arXiv preprint arXiv:2003.03415 (2020).

[47] H. Touchette, Physics Reports **478**, 1 (2009).