Extremely large scale simulation of a Kardar-Parisi-Zhang model using graphics cards

Jefrrey Kelling (1) and Géza Ódor (2)
(1) Institute of Ion Beam Physics and Materials Research
Helmholtz-Zentrum Dresden-Rossendorf
P.O. Box 51 01 19, 01314 Dresden, Germany
(2) Research Institute for Technical Physics and Materials Science,
P.O. Box 49, H-1525 Budapest, Hungary

The octahedron model introduced recently has been implemented onto graphics cards, which permits extremely large scale simulations via binary lattice gases and bit coded algorithms. We confirm scaling behavior belonging to the 2d Kardar-Parisi-Zhang universality class and find a surface growth exponent: $\beta = 0.2415(15)$ on $2^{17} \times 2^{17}$ systems, ruling out $\beta = 1/4$ suggested by field theory. The maximum speedup with respect to a single CPU is 240. The steady state has been analyzed by finite size scaling and a growth exponent $\alpha = 0.393(4)$ is found. Correction to scaling exponents are computed and the power-spectrum density of the steady state is determined.

We calculate the universal scaling functions, cumulants and show that the limit distribution can be obtained by the sizes considered. We provide numerical fitting for the small and large tail behavior of the steady state scaling function of the interface width.

PACS numbers: 05.70.Ln, 05.70.Np, 82.20.Wt

I. INTRODUCTION

The research of the nonlinear stochastic differential equation and universality class introduced by Kardar, Parisi and Zhang (KPZ) [1] is in the forefront of interest nowadays again [2,3]. This is largely due to the progress in exact solutions for various one-dimensional realizations and initial conditions (see for example [4–8]). This equation can describe the dynamics of simple growth processes in the thermodynamic limit [9,10], randomly stirred fluid [11], directed polymers in random media [12] dissipative transport [13,14], and the magnetic flux lines in superconductors [15]. Due to the mapping onto the Asymmetric Exclusion Process (ASEP) [16] in one dimension it is also a fundamental model of nonequilibrium particle system [17]. The KPZ equation specifies the evolution of the height function $h(x,t)$ in the $d$ dimensional space

$$\partial_t h(x,t) = v + \sigma_2 \nabla^2 h(x,t) + \lambda (\nabla h(x,t))^2 + \eta(x,t).$$

Here $v$ and $\lambda$ are the amplitudes of the mean and local growth velocity, $\sigma_2$ is a smoothing surface tension coefficient and $\eta$ roughens the surface by a zero-average, Gaussian noise field exhibiting the variance $\langle \eta(x,t)\eta(x',t') \rangle = 2D\delta^d(x-x')(t-t')$. The letter $D$ denotes the noise amplitude and $\langle \rangle$ means the distribution average. The equation is solvable in $(1+1)d$ due to the Galilean symmetry $\mathcal{G}$, and an incidentual fluctuation-dissipation symmetry $\mathcal{F}$, while in higher dimensions approximations are available only. The model exhibits diverging correlation length, thus scale a invariance, that can be understood by the steady current in the ASEP model corresponding to the up-down anisotropy of KPZ. Therefore KPZ equation has been investigated by renormalization techniques [18–20]. As the result of the competition of roughening and smoothing terms, models described by the KPZ equation exhibit a roughening phase transition between a weak-coupling regime ($\lambda < \lambda_c$) and a strong coupling phase. The strong coupling fixed point is inaccessible by perturbative renormalization group (RG) method. Therefore, the KPZ phase space has been the subject of controversies for a long time [21,22] and the strong coupling fixed point has been located by non-perturbative RG very recently [23].

Discretized versions of KPZ have also been studied a lot [24,27], for a review see [9]). Recently we have shown [28,29] the mapping between the KPZ surface and the ASEP [30,31] can straightforwardly be extended to higher dimensions. In two dimensions the mapping is just the simple extension of the rooftop model to the octahedron model as can be seen on Fig. 2 of [28]. The surface built up from the octahedra can be described by the edges meeting in the up/down middle vertexes. The up edges in the $x$ or $y$ directions are represented by $\sigma_{x/y} = +1$-s, while the down ones by $\sigma_{x/y} = -1$ in the model. This can also be understood as a special 2$d$ cellular automaton [32] with the generalized Kawasaki updating rules

$$\left( \begin{array}{cc} -1 & 1 \\ 1 & -1 \end{array} \right) \frac{p}{q} \left( \begin{array}{cc} 1 & -1 \\ -1 & 1 \end{array} \right)$$

with probability $p$ for attachment and probability $q$ for detachment. We have confirmed that this mapping, using the parametrization: $\lambda = 2p/(p + q) - 1$, reproduces the one-point functions of the continuum model. This kind of generalization of the ASEP model can be regarded as the simplest candidate for studying KPZ in $d > 1$: a one-dimensional model of self-reconstructing $d$-mers [33] diffusing in the $d$-dimensional space. Furthermore this lattice gas can be studied by very efficient simulation...
methods.

Now we implement dynamic, bit-coded simulations of the conserved lattice gas models for graphics cards (GPUs), allowing very large system sizes $L \times L$. The surface heights are reconstructed from the slopes

$$h_{i,j} = \sum_{l=1}^{i} \sigma_x(l,1) + \sum_{k=1}^{j} \sigma_y(i,k)$$

and the squared interface width

$$W^2(L,t) = \frac{1}{L^2} \sum_{i,j} h_{i,j}^2(t) - \left( \frac{1}{L} \sum_{i,j} h_{i,j}(t) \right)^2.$$  

was calculated at certain sampling times ($t$). In the absence of any characteristic length, growth processes are expected to follow power-law behavior and the surface can be described by the Family-Vicsek scaling law:

$$W(L,t) \simeq L^{\alpha} f(t/L^z),$$

with the universal scaling function $f(u)$

$$f(u) \sim \begin{cases} u^\beta & \text{if } u \ll 1 \\ \text{const.} & \text{if } u \gg 1 \end{cases}$$

Here $\alpha$ is the roughness exponent of the stationary regime, when the correlation length has exceeded the system size $L$; and $\beta$ is the growth exponent, describing the intermediate time behavior. The dynamical exponent $z$ is just the ratio

$$z = \alpha/\beta.$$  

II. BIT-CODED GPU SIMULATIONS

The height of each surface site is thoroughly determined by two slopes, along the $x$ and $y$ axes respectively, whose absolute values are restricted to unity. Thus at each site two bits of information are required, hence a chunk of $4 \times 4$ sites is encoded in one 32-bit word.

Two slightly different layers of parallelization are used that reflect the two layered compute architecture provided by GPUs [35]: not communicating blocks at device level and communicating threads at block level. Both layers use domain decomposition with dead borders, i.e. conflicts at the subsystem borders are avoided simply by not updating them [see Fig. 1]. A random translation is applied to the origin of the decomposition periodically to preserve statistics. To avoid having to deal with non-32-bit aligned memory these translations are restricted to multiples of four sites.

The complete system is stored in global device memory, each block cell is copied into the block-local shared memory for precessing. Thus moving the origin of the device level decomposition results in cutting out the proper piece of the system taking the periodic boundary conditions into account. Moving the origin at each Monte Carlo step (MCS), i.e. by one overall update of the system, proved to be sufficient and the results are undistorted.

The size of a thread cell is set to be $8 \times 8$ sites, the smallest possible choice, to maximize the number of threads per block. Due to this small subsystem size a new origin for the block level decomposition is picked at each update attempt, thus 64 times per MCS. Additionally the borders are not dead but delayed, i.e. when a thread picks the border of its cell to change it waits for the threads updating bulk sites to finish their updates. Corner sites are further delayed.

If a thread hits a site belonging to a block border it does nothing, slightly reducing the actual number of update attempts per MCS. This is a minor effect, the ratio of block border to overall system size is approximately $1/128$, and impacts the pre-factors of the scaling, but not exponents.

For random number generation each thread uses a 32-bit, linear congruential random number generator (LCRNG) with different seed [36]. Similar generators were previously successfully applied in GPU simulations of the Ising model [37]. Depending on the system size the generators have to be periodically reseeded, which potentially introduces the same correlations as using multiple generators in parallel. However, since no deviations from the earlier CPU results have been observed, we assume this to not disturb the statistics. Correlations resulting from parallel usage could only have local effects on the updates of a block cell. Moving the origin of the block level decomposition should effectively destroy such correlations. By the same argument reseeding the genera-
tors, using a Mersenne Twister generator \[38\] running on the CPU, has no negative effect at all. Part of the results were double-checked, employing a skip-ahead 64-bit LCRNG \[39\] with no need for reseeding.

The simulations were performed on a C2070 GPU with 6GB graphics memory, which allowed for a maximum system size: $2^{17} \times 2^{17}$ (4GB of memory required). Figure 2 shows benchmark results for the simulation.

![FIG. 2: (Color online) Time in seconds for one MCS on a C2070 (black bullets, left axis). Speedup over a non parallel implementation on a Core i5 750@3GHz (red squares, right axis). The jump in speedup between $\log_2 L = 12$ and $13$ results from the system exceeding the CPUs L3 cache at this point. The maximum speedup factor achieved was roughly 240.](image)

The benchmarks only consider bare simulation times and exclude the time needed to transfer data between host and device. For large sizes, where the system exceeds the CPUs L3 cache, the performance drops significantly, this could be avoided in a CPU implementation using domain decomposition designed to optimize memory access of the CPU cache.

### III. SURFACE GROWTH SCALING

We have run $10 - 1000$ independent simulations for sizes: $L = 2^8, 2^9, 2^{10}, 2^{11}, 2^{12}, 2^{13}, 2^{16}, 2^{17}$ by starting from half filled (striped) lattice gases. This causes an intrinsic width of the initial zig-zag surface state, with $W^2(L,0) = 1/4$, a leading order constant correction to scaling, that we subtract at the beginning of the scaling analysis. The time between measurements increases exponentially

$$t_{i+1} = (t_i + 10) \cdot e^m,$$

with $m > 0$, $t_0 = 0$, \(8\)

where the program calculates and writes out the width of the surface. We used $m = 0.01$ to study the growth in larger systems, and $m = 0.001$ or $m = 0.0001$ to collect more data points in the steady state. By the scaling analysis we disregarded the initial time region: $t < t_{\text{mic}} \approx 50$, when basically an uncorrelated, random deposition process goes on. The growth is expected to follow simple scaling \(9\) asymptotically and we assume corrections in the form

$$W(t, L \to \infty) = bt^\beta (1 + b_0 t^\phi_0 + b_1 t^\phi_1 + ...). \quad \text{(9)}$$

For finite system, when the correlation length exceeds $L$, the growth crosses over to a saturation, with the expected scaling behavior

$$W(t \to \infty, L) = aL^\alpha (1 + a_0 L^\omega_0 + a_1 L^\omega_1 + ...). \quad \text{(10)}$$

To see the corrections clearly we determined the effective exponent of $\beta$, as the discretized, logarithmic derivative of $11$

$$\beta_{\text{eff}}(t) = \frac{\ln W(t, L \to \infty) - \ln W(t', L \to \infty)}{\ln(t) - \ln(t')} \quad \text{(11)}$$

using $t/t' = 2$. As Fig. 3 shows the $\beta_{\text{eff}}(t)$ curves converge to the same asymptotic value for different sizes as $1/t \to 0$, albeit for smaller systems the fluctuations are larger and the scaling breaks down as $\xi(t) \approx L$. One can read-off the most precise $\beta = 0.2415(15)$ estimate from the largest system ($L = 2^{17}$), where the simulations were followed up to $t = 70,000$ MCS. This agrees with our former estimates for this model \[28, 29\], but now the error margin is sufficiently small to exclude a convergence to the field theoretical value $\beta = 1/4 \ [10]$ via the analytic corrections \[10\].

Following the subtraction of the constant leading-order term, corresponding to $\phi_0 = -\beta$, $b_0 = 1/2$, the remaining corrections are seemingly small and the oscillations hinder to fit them out very precisely. We determined the next leading order correction exponent by fitting with the from \[12\]

$$\beta_{\text{eff}}(t) = \beta + b_1 \phi_1 t^\phi_1, \quad \text{(12)}$$

in the time window $t > t_{\text{mic}}$ and before the saturation region. From the largest system fit we obtained: $\phi_1 = 1.05$ and $b_1 = -0.12$.

On Fig. 3 we can observe that the local slopes do not change for late times, therefore assuming a $W \propto t^{0.2415}$ asymptotic scaling we determined the probability distribution around this law for the largest size: $L^{17}$ for $1300 < t < 70000$ MCS. We calculated the distribution of $y = W(t)/t^{0.2415}$ as shown on the left inset of Fig. 3. This opens up the possibility for a future comparison with a solution like in one dimension \[11\].

Next we investigated the scaling in the steady state. This could be achieved in smaller systems with a higher data sampling resolution. We confirmed that the data corresponds to the steady state by visual inspection of the $W^2(t,L)$ as well as by analyzing the $P(W^2)$ distribution. Similarly to the time dependence we determined the effective exponent of the roughness, which can be defined as the logarithmic derivative of the width

$$\alpha_{\text{eff}}(L) = \frac{\ln W(t \to \infty, L) - \ln W(t \to \infty, L/2)}{\ln(L) - \ln(L/2)} \quad \text{(13)}$$
Finite size scaling was done for systems of linear sizes in between $L_{\text{min}} = 2^8$ and $L_{\text{max}} = 2^{13}$ and by considering the corrections using the fitting form

$$\alpha_{\text{eff}}(L) = \alpha + a_1 \omega_1 L^{\omega_1}. \quad (14)$$

The local slopes of the steady state values $\alpha_{\text{eff}}(1/L)$ are shown on Fig. 4. The fitting results in: $\alpha = 0.393(4)$, $a_1 = -1.24$ and $\omega_1 = 1.16$. Using the $\alpha = 0.393(3)$ and $\beta = 0.2415(15)$ estimates the dynamical scaling exponent is $z = \alpha/\beta = 1.627(26)$. With these values we get $\alpha + z = 2.02(3)$, which satisfies the scaling law expected by the Galilean symmetry. Fig. 5 shows a perfect data collapse with these exponents over several decades.

We also investigated the power spectrum density (PSD) of the interface

$$S(k, t) = \langle h(k, t) h(-k, t) \rangle, \quad (15)$$

where the height in the Fourier space is computed as

$$h(k, t) = \frac{1}{L^{d/2}} \sum_{\chi} [h(x, t) - \langle h \rangle] \exp(i k \chi). \quad (16)$$

We computed $h(k, t)$ from the surface profiles with the FFT method and determined $S(k, t)$ by averaging over $x$ and $y$ directions. In the steady state the PSD is expected to scale as $S \sim k^{-z-2\alpha}$, which can be confirmed for $0.002 < k < 0.1$ (see inset of Fig. 5). For larger $k$ values we can see a growth of the $S(k)$ function, which is the consequence of the lattice regularization.

### IV. PROBABILITY DISTRIBUTIONS

The exact form of the spatially averaged height distribution $P(h)$ of the KPZ model in one dimension is a hot topic of statistical physics [48] and provides a definition of the KPZ universality class. The $P_L(W^2)$ distribution in the steady state is also known exactly; in closed form for small and large $x$ asymptotically [42]. In two dimensions not much is known about this distribution.

In [22, 42] it was shown that the width distributions $\Psi_L(x) = \langle W^2 \rangle P_L(W^2/(W^2))$ of discrete KPZ surfaces exhibit universal behavior. Now we determined the probability distributions $P_L(W^2)$ and calculated $\Psi_L(x)$ for $L = 2^8, 2^9, 2^{10}, 2^{11}, 2^{12}, 2^{13}$ with the GPU code by measuring $W^2$ in the steady state. Averaging was done over $N = 20 - 100$ runs and $10^4 - 10^5$ time steps. In case of the largest, $2^{13}$ case the steady state averaging was done between $5 \times 10^7$ and $10^8$ MCS. As Fig. 4 shows the data collapse is very good around $x = 1$, but deviations occur in the large and small $x$ asymptotics due to the lack of sample points there. It is very difficult to collect enough statistics for the extremal cases as the width of $P_L(W^2)$ grows as $L^{2\alpha}$.
By studying the finite size effects of extreme value
statistics it was discovered \cite{13} that there is also
universality in the first order (shape) correction to the
limit distributions. It was also shown that if the finite size
corrections of the cumulants can be neglected the shape
corrections can be expressed via the limit distributions.
To see this let us write $P_L(W^2)$ in terms of the cumulant
generating function
\begin{equation}
P_L(W^2) = \int \frac{dq}{2\pi} \exp \left[ -iq(W^2 - \kappa_1) + \sum_{n=2}^{\infty} \frac{(iq)^n}{n!} \kappa_n \right],
\end{equation}
where $\kappa_n$-s are the $n$-th cumulants of $W^2$ (i. e. $(W^2)$
= $\kappa_1 = \nu_1$), related to the $n$-th, non-central moments ($\nu_n$)
as:
\begin{align}
\kappa_1 &= \nu_1, \\
\kappa_2 &= \nu_2 - \nu_1^2, \\
\kappa_3 &= \nu_3 - 3\nu_1\nu_2 + 2\nu_1^3, \\
\kappa_4 &= \nu_4 - 4\nu_3\nu_1 - 3\nu_2^2 + 12\nu_1^2\nu_2 - 6\nu_1^4, \\
&\quad \text{...}
\end{align}
\begin{equation}
\kappa_n \propto L^{2n\alpha}
\end{equation}
Let us assume that the corrections to scaling of the cumulants are of the form
\begin{equation}
\kappa_n = L^{2n\alpha}(\kappa_n^0 + \kappa_n^1 L^{-\omega_1} + \ldots)
\end{equation}
To check this we determined the $n = 1, 2, 3, 4$ cumulants
from $W^2$ data and performed a finite size scaling analysis. The corrections to scaling (11) were found to be
negligible, as shown on Fig. 7, hence the universal limit
distribution in principle can very well be approximated
from the finite $L$ results.

Finally we tried to fit the small and large $x$ asymptotics
of $\Psi(x)$ with similar functional forms that is known ex-
actly in one dimension \cite{42}. This assumption is based on
the similarity of the underlying model, i.e. directed mi-
gration of dimers instead of particles. When we applied
for the small $x$ ($x < 0.75$) part of the $\Psi_L(x)$ the general
form
\begin{equation}
x^A(B-x) \exp \left( C/x^D \right)
\end{equation}
we found stable nonlinear fittings with $C \simeq 2$ and $D \simeq 2$
in contrast to one dimension, where $D = 1$. This is similar to the small $x$ extreme value statistics of the $1/f^\alpha$
noise, where one obtains $\exp(-a/x^\beta)$ with $\beta$ depending
on $\alpha$. We fixed $C = 2$, $D = 2$ and tried to get a general
form with integer coefficients up to the second order. We
obtained
\begin{equation}
x^{-8}(10 - x) \exp \left(-2/x^2\right) \left(1 + x^{-38}(9 - x) \exp(-9.25/x^2)\right)
\end{equation}
in good agreement with the $L = 2048$ data as shown on
Fig. 8.

For the larger $x$ part we assumed again the form of the
one dimensional model
\begin{equation}
E \exp(Fx)/x
\end{equation}
and obtained a nice agreement with: $E = 2.915$, $F = -2.572$. The least squares fit error was smaller that by a
stretched exponential ansatz.

\section{Conclusions and discussion}

In conclusion we have developed a bit-coded CUDA
program, which simulates very efficiently a 2 + 1 dimen-
sional discrete KPZ growth model (the octahedron model
\cite{28}) via binary lattice gases. Using this tool we could
achieve extraordinary large sizes up to $2^{17} \times 2^{17}$ with a
maximum speedup 240 on NVIDIA Fermi Cards with respect to a single 3 GHz CPU core. This allows us to resolve debates over the scaling exponents by performing very precise scaling analysis of the interface width.

Our growth exponent estimate \( \beta = 0.2415(15) \) is somewhat bigger than the results of \([44]\) \((\beta = 0.221(2)), \[45]\) \((\beta = 0.229(5))\) and \([46]\) \((\beta = 0.240(1))\). It matches our former estimates for this model 0.245(5) \([29]\), but excludes definitely the \( \beta = 0.25 \) field theoretical result. We also estimated the leading order correction to scaling exponent: \( \phi_1 = 1.05 \).

The independent roughness exponent result \( \alpha = 0.393(3) \) is in the middle of the range obtained by various numerical exponent estimates: i.e. between \( \alpha = 0.36 \[44,47\], \alpha = 0.385(5) \[40\] \) and the \( \alpha = 0.4 \) field theoretical result, well out of error margin. This agrees with our former \( \alpha = 0.395(5) \[29\] \) and with the simulation results of \([21]\) \((\alpha = 0.393(3))\). In the latter case even the correction to scaling exponent \( \omega_1 = 1.16 \) is the same.

We analyzed the surface in the steady state by the power spectrum density method and confirmed the KPZ scaling for several decades above the lattice cut-off. We determined the universal scaling function and the cumulants of the surface width for different sizes and obtained the limit distribution via correction to scaling analysis. We gave analytical fitting for the small and large asymptotics. As compared to one dimension \([42]\), where a linear \( x \) dependence in the exponential is known exactly, we found \( x^2 \) dependence for small \( x \). For the large \( x \) deviations the \( \exp(Fx)/x \) tail fits better than a stretched exponential functions as suggested in \([45]\).

Our model and code proves to be a very efficient tool to study not only the \( (2+1) \) dimensional KPZ and ASEP models but, more generally it can be used in the research of fundamental nonequilibrium thermodynamical quantities like the large deviation function or entropy production \([49,50]\). It is also straightforward to extend it to study more complex system exhibiting pattern formation \([51,52]\), the effect of quenched disorder \([53]\), the time-dependent structure factor \([54]\) or to higher dimensions \([29]\).

One may also ask if the results for this discrete model describe those of the continuum KPZ equation. In fact this is not an obvious question in \( d > 1 \) dimensions \([55]\). However, we think that the irrelevancy of anisotropy by renormalization group studies \([50]\) excludes this in two dimensions and we find \( (2+1) \) dimensional KPZ universality class behavior.

On completion of this study we discovered another way of accelerating our algorithm, which provides an additional factor of \( \sim 1.8 \) with respect to the simulations presented here. The technical details will published elsewhere \([57]\).

Acknowledgments:

We thank Zoltan Rácz and Karl-Heinz Heinig for the useful discussions and E. Katzav for his comments. Support from the Hungarian research fund OTKA (Grant No. T077629), the bilateral German-Hungarian exchange program DAAD-MÖB (Grant Nos. 50450744, P-MÖB/854) and OSIRIS FP7 is acknowledged. The authors thank NVIDIA for supporting the project with high-performance graphics cards within the framework of Professor Partnership and the computer center of HZDR for the GPU cluster access.

---