

AN ANALYSIS OF 2D PERCOLATION

ALAN SAMMARONE



An exploration of percolation 2D theory

Institut für Theoretische Physik
Universität Leipzig

May 2020

ABSTRACT

The current state of percolation theory is hard to assess. Research often uses different methods, language and notation, is hard to reproduce, and sometimes even provides conflicting results. We aim to give a numerically detailed, reproducible overview of the classic 2D percolation problem, along with data and open source code that allow the reader to easily reproduce the results presented, as well as present a simple numerical analysis of the relatively uncommon model called Mandelbrot percolation. We hope to facilitate further research, as well as provide a starting point for anyone interested in the topic.

*We have seen that computer programming is an art,
because it applies accumulated knowledge to the world,
because it requires skill and ingenuity, and especially
because it produces objects of beauty.*

— **knuth:1974** [knuth:1974]

ACKNOWLEDGEMENTS

CONTENTS

I	CLASSIC 2D PERCOLATION	1
1	THEORETICAL ASPECTS	3
1.1	Setup	3
1.2	Percolation	3
1.3	Percolating cluster strength	5
1.4	Mean cluster size	5
1.5	Correlation function and correlation length	6
1.6	Cluster size distribution	7
2	EXPERIMENTAL METHODS AND RESULTS	9
2.1	2D lattice simulation	9
2.2	Percolation probability	9
2.3	Percolating cluster strength	13
2.4	Mean cluster size	14
2.5	Correlation function and correlation length	15
II	MANDELBROT PERCOLATION	17
III	APPENDIX	19
A	CODE	21
B	DATA	23
C	TABLES	25

LIST OF FIGURES

Figure 1	A 1D lattice of size L , with a few dead nodes .	4
Figure 2	A finite piece of an infinite 2D lattice. The largest cluster is made up of the nodes with circles painted, while the smaller clusters are denoted by open circles.	4
Figure 3	Percolation probability curve for various lattice sizes	6
Figure 4	Percolation probability curve for various lattice sizes	10
Figure 5	Percolation probability curve zoomed in, along with the hyperbolic tangent least squares fit. .	11
Figure 6	Finite-size scaling analysis of the hyperbolic tangent function	12
Figure 7	Percolating cluster strength versus occupation probability for various lattice sizes	13
Figure 8	Mean finite cluster size versus occupation probability for various lattice sizes	14
Figure 9	Correlation function for various sizes and occupation probabilities	15
Figure 10	Correlation length for multiple lattice sizes . .	15

LIST OF TABLES

Table 1	Hyperbolic tangent least squares fit parameters	25
---------	---	--------------------

LISTINGS

ACRONYMS

PMF Probability mass function

Part I

CLASSIC 2D PERCOLATION

THEORETICAL ASPECTS

1.1 SETUP

A lattice (G, N, S) is set of nodes G , a neighbour relation N from G to the power set of G : $N: G \rightarrow \mathcal{P}(G)$, and a state map $S: G \rightarrow Q \subseteq \mathbb{N}$. The precise nature of G and N depends on exactly what type of lattice we're considering. For simplicity, our analysis and results here will be limited to a 2D square lattice, which can be either infinite or finite. In this case, G is a subset of the set of ordered pairs of natural numbers:

$$G \subseteq \mathbb{N} \times \mathbb{N} = \{(i, j) \mid i, j \in \mathbb{N}\} \quad (1)$$

So, for example, $(3, 2)$ and $(8843, 0)$ could be elements of G . If we're considering an infinite lattice, then we have

$$G = \mathbb{N} \times \mathbb{N}$$

N specifies a set of neighbours for each element of G . If one views the lattice as a graph, then N contains information about the edges of the graph. In the (infinite) 2D square lattice case, we have

$$\begin{aligned} N: G &\rightarrow \mathcal{P}(G) \\ (i, j) &\mapsto \{(i-1, j), (i+1, j), (i, j-1), (i, j+1)\} \end{aligned} \quad (2)$$

For every node in G , this map assigns 4 neighbours. If one thinks of the lattice as a grid, then for every node (i, j) this map assigns the left, right, top and bottom nodes as its neighbours.

S simply assigns a "state" in the set Q to each node in G . In standard percolation theory, each node can be in one of two states, so $Q = \{0, 1\}$. These two states are also called dead/alive, black/white or active/inactive.

A cluster C in the lattice is defined as a subset of connected nodes (i, j) in the lattice with such that $S(i, j) = 1$. Connected here means that for every pair $(i_1, j_1) \in C$ and $(i_2, j_2) \in C$ there exists a sequence of neighbouring nodes with (i_1, j_1) being the first item in the sequence and (i_2, j_2) being the last.

1.2 PERCOLATION

Percolation theory refers to the study of the structure of clusters formed in the lattice, given some choice of state map S . In standard percolation theory, one randomly (with probability p) assigns either

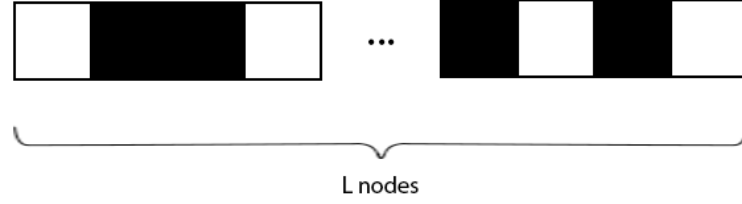


Figure 1: A 1D lattice of size L , with a few dead nodes

state 0 or state 1 to each and every node in the lattice. Under these circumstances, the system presents a phase transition, when one varies the occupation probability p . There exists a critical value p_c such that for $p > p_c$, the infinite lattice is said to "percolate", i.e. clusters spanning the whole lattice appear. For the 2D square lattice, the threshold is $p_c = 0.59274621(13)$. This number changes for different choices of G and N .

[Figure 1](#) and [Figure 2](#) provide examples of percolation in 1D and 2D lattices.

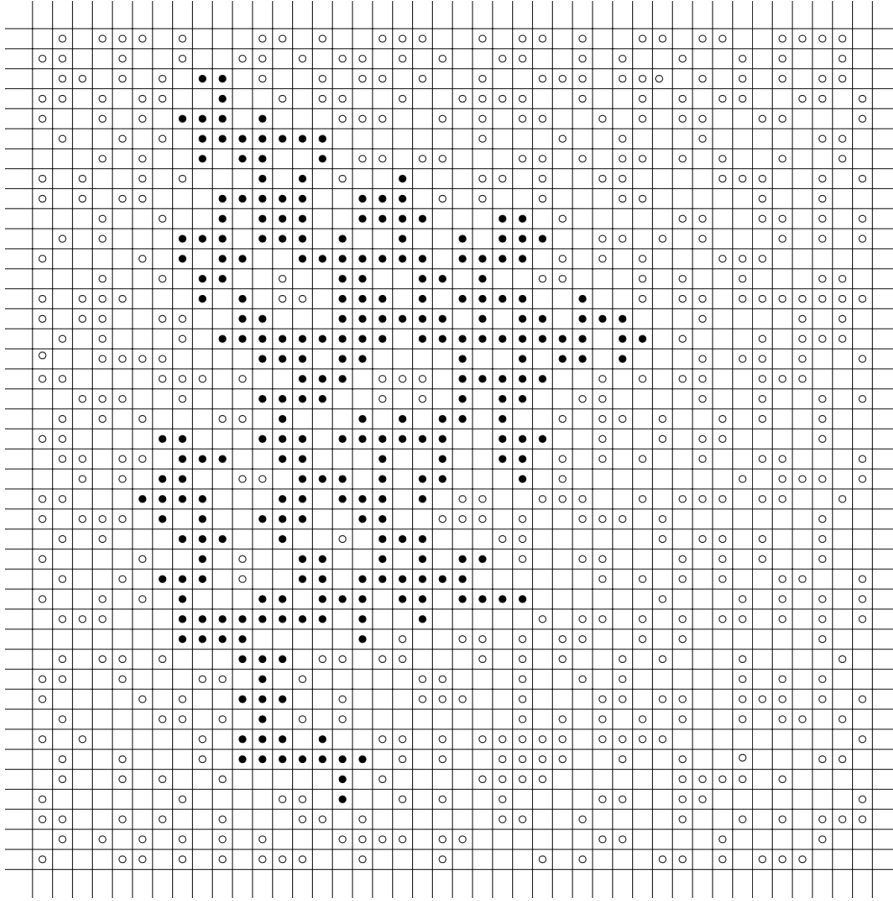


Figure 2: A finite piece of an infinite 2D lattice. The largest cluster is made up of the nodes with circles painted, while the smaller clusters are denoted by open circles.

In biology, percolation theory has been used to successfully predict the fragmentation of biological virus shells (REF), with the percolation threshold of Hepatitis B virus capsid predicted and detected experimentally (REF). In environmental science, percolation theory has been applied to studies of how environment fragmentation impacts animal habitats (REF) and models of how the plague bacterium *Yersinia pestis* spreads (REF).

Our goal will be to study the behavior of some quantities of interest associated with percolation models.

1.3 PERCOLATING CLUSTER STRENGTH

The first interesting quantity we can look at is the percolating strength $P(p)$. This quantity represents the probability that a randomly picked node belongs to the percolating cluster. Another way to think of this quantity is that it represents the fraction of the whole lattice that is covered by the percolating cluster.

For small values of p , all islands are finite. Starting at any randomly site, it is impossible to get arbitrarily far away by walking only along the connected sites. As we increase p , the largest cluster becomes infinite at a precise value $p = p_c$. When this happens, it is possible to get arbitrarily far from a starting point in the infinite cluster by walking along the connected sites. For $p < p_c$, there are no infinite clusters, so $P(p)$ vanishes. However, for $p > p_c$, $P(p)$ monotonically increases. For $1 \gg (p - p_c) > 0$, that is, for p greater but close to p_c , we have (REF 1):

$$P(p) \propto (p - p_c)^\beta$$

The qualitative behaviour of $P(p)$ can be seen in figure (FIG 1)

β is a universal critical exponent which depends only on the lattice dimension, and not on its geometry. In 2 dimensions, $\beta = \frac{5}{36}$, is the same for the square, hexagonal, or any other geometry (REF 1).

$P(p)$ is one of the order parameters for percolating systems, since it is zero for $p < p_c$, and non-zero for $p > p_c$.

1.4 MEAN CLUSTER SIZE

Another interesting quantity is $\chi(p)$, the mean cluster size, which is the average number of nodes in among finite clusters. One can think of each node in the cluster as having one "unit of mass", in which case the mean cluster size represents the average mass. Just like $P(p)$, for $\chi(p)$, we have

$$\chi(p) \propto |p - p_c|^{-\gamma}, \quad \text{for } |p - p_c| \ll 1$$

The qualitative behaviour of $\chi(p)$ can be seen in [Figure 3](#).

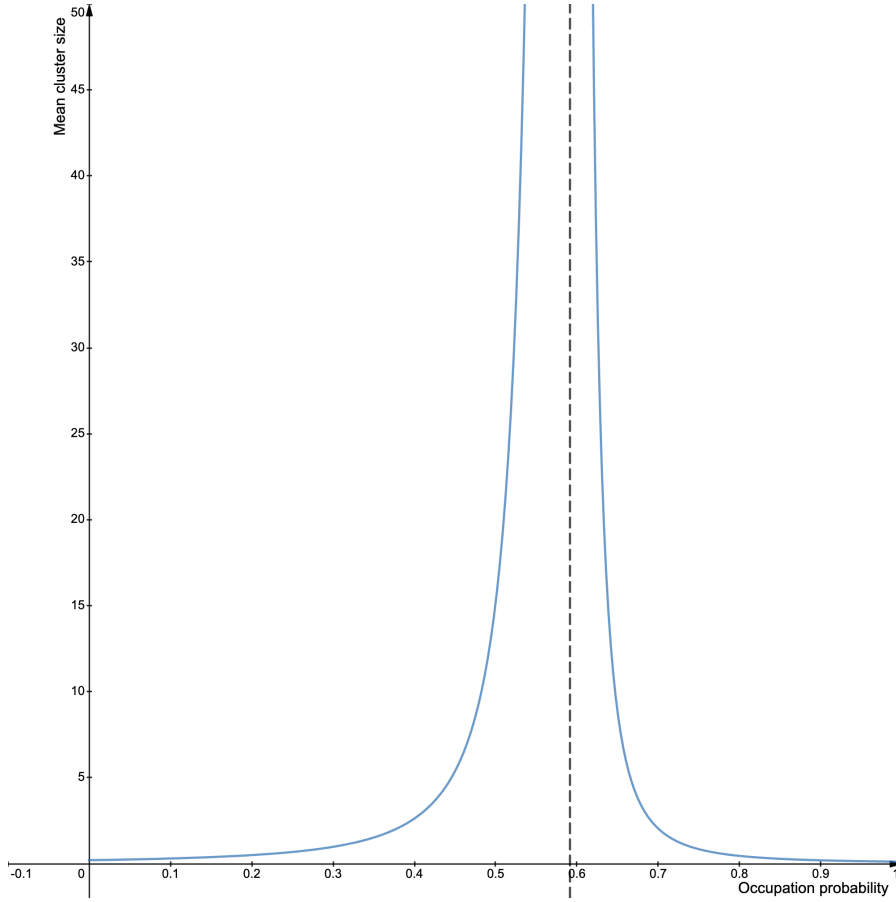


Figure 3: Percolation probability curve for various lattice sizes

Again, analogously to the percolating cluster strength $P(p)$, γ is a universal critical exponent, in the sense that it's independent of the local geometry, and depends only on the dimension of the lattice. In two dimensions, $\gamma = \frac{43}{18}$. One important difference between $\chi(p)$ and $P(p)$, however, is that $\chi(p)$ scales as $|p - p_c|^\gamma$ for both $p < p_c$ and $p > p_c$. The constants of proportionality are different in either side.

1.5 CORRELATION FUNCTION AND CORRELATION LENGTH

The correlation function $g(\vec{r})$ encodes how likely it is that two clusters separated by a displacement vector \vec{r} are part of the same (finite) cluster. In 2D, based on rotational symmetry and in order to simplify the analysis, one often ignores the vector character of \vec{r} and studies only $g(r)$, where $r = |\vec{r}|$.

For large values of r and $p \neq p_c$, the correlation function decays exponentially:

$$g(r) \propto e^{-\frac{r}{\xi(r)}}, \quad \text{for } p \neq p_c, \quad r \gg 1$$

$\xi(r)$ is the so called correlation length. One interpretation is that it measures the size of correlations in the lattice, that is, the typical

diameter of an island. With this interpretation, one does not need to consider distances larger than ξ : a finite piece of the lattice, larger than ξ , presents the same behavior as an infinite lattice.

This exponential form is only the leading factor in $g(r)$. At $p = p_c$, long range correlations appear, $\xi(p)$ diverges, and a power law leading factor appears:

$$g(r) \propto r^{-\eta}, \quad \text{for } p = p_c, \quad r \gg 1$$

Once more, the critical exponent η is universal. In 2D, $\eta = \frac{5}{24}$.

When $p \neq p_c$, the correlation length defines a characteristic length scale. This is not the case when $p = p_c$: there's no characteristic length scale. The lattice is self similar and any finite piece is not enough to describe the behaviour of the system.

Near the threshold p_c , the correlation length, again, behaves like

$$\xi(p) \propto |p - p_c|^{-\nu}, \quad \text{for } |p - p_c| \ll 1$$

with ν the corresponding critical exponent, also universal. In 2D, $\nu = \frac{4}{3}$.

It is interesting to notice here the importance of considering only finite clusters: the correlation length remains finite both below and above the threshold p_c . As we increase p past p_c , finite clusters begin forming connections to the infinite cluster, and so the probability that two nodes belong to the same finite cluster (which is what $g(r)$ measures) decreases, and so does the correlation length.

1.6 CLUSTER SIZE DISTRIBUTION

Perhaps the most important quantity we'll look at is the cluster size distribution, $n(s, p)$. This function describes the s -cluster density at occupation probability p , that is, the number of clusters with size s and probability p divided by the size of the lattice (in the case of infinite lattices, this quantity can be thought of either as an average over many large finite pieces of the infinite lattice, or as a limit as $L \rightarrow \infty$). At p_c , the cluster size follow a power law distribution:

$$n(s, p_c) \propto s^{-\tau}$$

where τ is another universal exponent (in two dimensions, $\tau = \frac{187}{91}$). Once more, the system lacks a characteristic length scale. Near p_c and for large s , $n(s, p)$ is a generalized homogeneous function of $p - p_c$ and $\frac{1}{s}$.

A function of two variables $F(x, y)$ is said to be generalized homogeneous function iff

$$F(c^\alpha x, c^\beta y) = c F(x, y) \quad \forall (c, \alpha, \beta)$$

In the case of $n(s, p)$, we have

$$n\left(\frac{c^\alpha}{s}, (p - p_c)c^\beta\right) = cn\left(\frac{1}{s}, p - p_c\right)$$

By choosing $c = s^{\frac{1}{\alpha}}$, one obtains

$$n(1, (p - p_c)c^{\beta/\alpha}) = n((p - p_c)c^{\beta/\alpha}) = s^{\frac{1}{\alpha}}n\left(\frac{1}{s}, p - p_c\right)$$

which allows us to conclude that

$$n(1/s, p - p_c) = s^{-1/\alpha}n((p - p_c)c^{\beta/\alpha})$$

What this means is that by measuring $n(s, p)$ in units of $n(s, p_c)$, that is, by taking $\frac{n(s, p)}{n(s, p_c)}$ one does not need to consider it as a function of two variables p and s : it reduces to a function which depends only on the combination $(p - p_c)s^\sigma$ for some sigma. In 2D, $\sigma = \frac{36}{91}$, which is also a critical exponent.

One of the reasons why $n(s, p)$ is one of the most interesting quantities to look at is that the order parameter (the percolating cluster strength), the mean finite cluster size, the correlation length and other quantities can all be derived from it. For this reason, $n(s, p)$ is called a generating function.

The percolating cluster strength $P(p)$, for example, can be obtained from

$$P(p) = p - \sum_s s n(s, p) \quad (3)$$

$n(s, p)$ describes the s -cluster count divided by the size of the lattice. So $s n(s, p)$ can be viewed as the total area proportion covered by all clusters of size s . In other other, it represents the probability that by picking a node at random, it will be part of an s cluster. By summing over all (finite) values of s , we get the probability that a randomly picked node belongs to a finite cluster. Since p can be viewed as representing the probability of belonging to any cluster (since any alive cell will, necessarily, belong to a cluster, and any node belonging to a cluster must be alive, these are equivalent descriptions), the difference between p and $\sum_s s n(s, p)$ gives us the probability of belonging to an infinite cluster, which is how $P(p)$ is defined.

The mean cluster size $\chi(p)$ is related to $n(s, p)$ through

$$\chi(p) = \sum_s s^2 n(s, p) \quad (4)$$

As noted before, $s n(s, p)$ represents the probability that a randomly picked node belongs to a (finite) s -cluster. Therefore, $\sum_s s^2 n(s, p)$ represents the expected value of s .

EXPERIMENTAL METHODS AND RESULTS

To experimentally verify the results described above, we ran more than **3 billion simulations**, generating approximately **250GB** of data. The code that used to run the simulation and to extract the statistics is available in Github, see Appendix 1 for more details. The full data is also available, see Appendix 2.

2.1 2D LATTICE SIMULATION

We used a $L \times L$ square, periodic lattice, which is topologically equivalent to a torus. Each node has four neighbours: one above it, one to the right, one below, and one to the left. The periodicity is represented by the fact that the right neighbors of the nodes in the rightmost column are the nodes in the leftmost column. Similarly, the left neighbours of the leftmost column are the nodes in the rightmost column, the top neighbours of the topmost row are the nodes in the bottom-most row, and the bottom neighbours of the bottom-most row are nodes in the topmost row.

Each node can be in one of two states: dead or alive. Initially, all nodes are in the alive state. With probability p , we change state of each individual node to dead. We then compute and store the set of clusters found in the lattice. A cluster is a set of connected nodes sharing the same state.

Each cluster can either be a finite or infinite. Since the lattice is periodic, a cluster is considered infinite if its 1D extension in either direction is greater than or equal to the size of the lattice L . Otherwise, it is a finite cluster. In other words, we imagine a bounding box around each cluster, and if at least of the dimensions of the box is greater than or equal to L , the cluster is infinite.

We generated around 3000 lattices for each pair (p, L) , with p ranging from 0 to 1 (with a higher density near the critical points) and L ranging from 16 to 1024.

2.2 PERCOLATION PROBABILITY

A lattice has percolated if it contains at least one infinite cluster. For each (p, L) , we count the number of percolating lattices and divide by the total number of lattices to compute the percolation probability.

In [Figure 4](#), we plot the percolation probability against the occupation probability p for various lattice sizes.

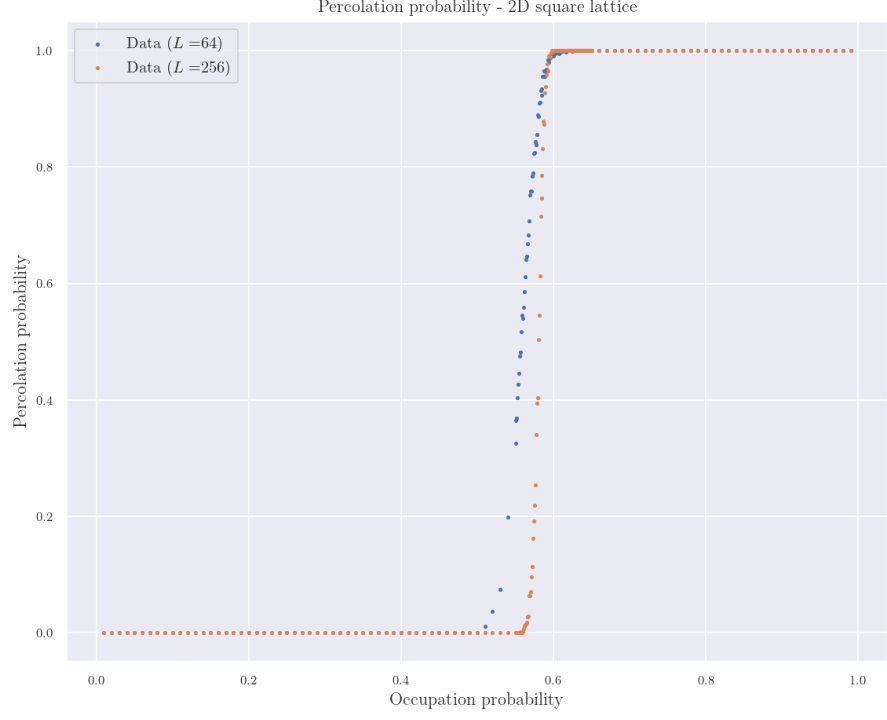


Figure 4: Percolation probability curve for various lattice sizes

By visual inspection, we can hypothesize that the curve is a hyperbolic tangent curve:

$$y(x) = \tanh(x)$$

We can shift and scale this function so that it's centered at $x = \beta$, and such that it's domain matches our data, i.e. $\mathcal{D}(y) = [0, 1]$. We'd also like a parameter that controls how steep the transition is between $y = 0$ and $y = 1$ - this can be done by scaling x by a multiplicative factor α . The function becomes

$$y(x) = \frac{\tanh(\alpha(x - \beta)) + 1}{2}$$

In [Figure 5](#), we can see the the fitted data points. [Table 1](#) contains the precise values.

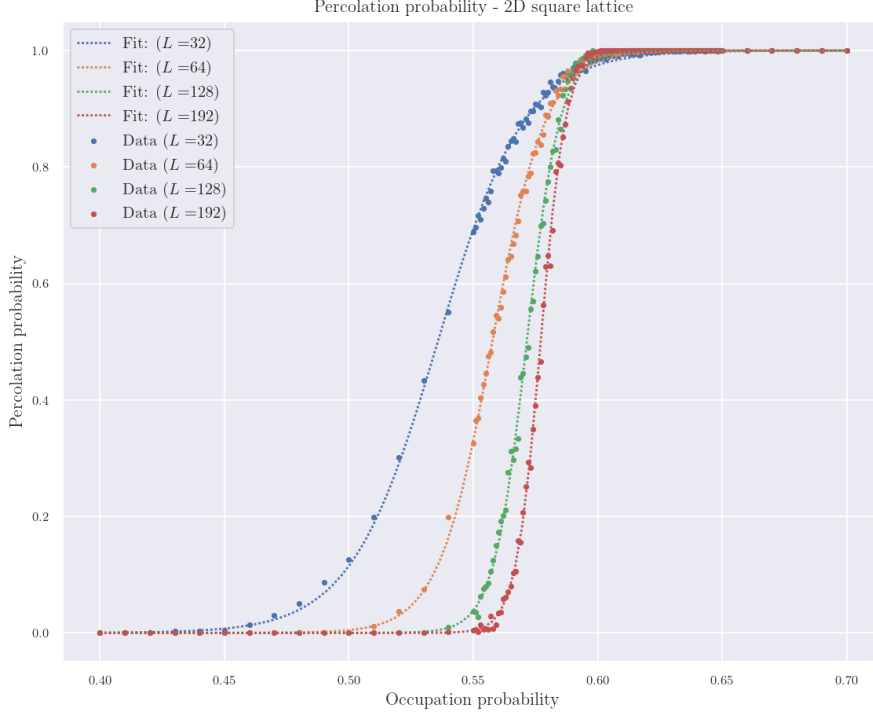


Figure 5: Percolation probability curve zoomed in, along with the hyperbolic tangent least squares fit.

We notice that the the bigger the lattice, the faster it jumps from $y = 0$ to $y = 1$. This is consistent with the widely known fact that as L increases, i.e. in the limit $L \rightarrow \infty$, the transition becomes a step function at $p = p_c$.

Another interesting fact to notice is that with the structure of the function we're fitting, β represents the point at which the percolation probability first reaches $\frac{1}{2}$.

These two observations suggest that studying the behaviour of α and β as a function of the lattice size L might be interesting. We can see the results obtained in [Figure 6](#), along with a so called finite-size scaling analysis.

Percolation probability - finite-size scaling

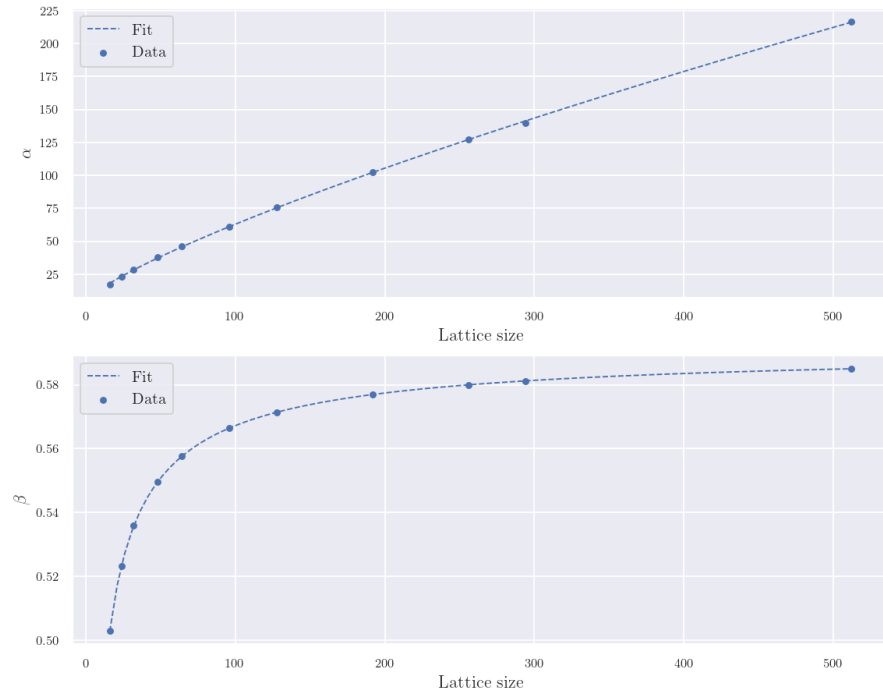


Figure 6: Finite-size scaling analysis of the hyperbolic tangent function

2.3 PERCOLATING CLUSTER STRENGTH

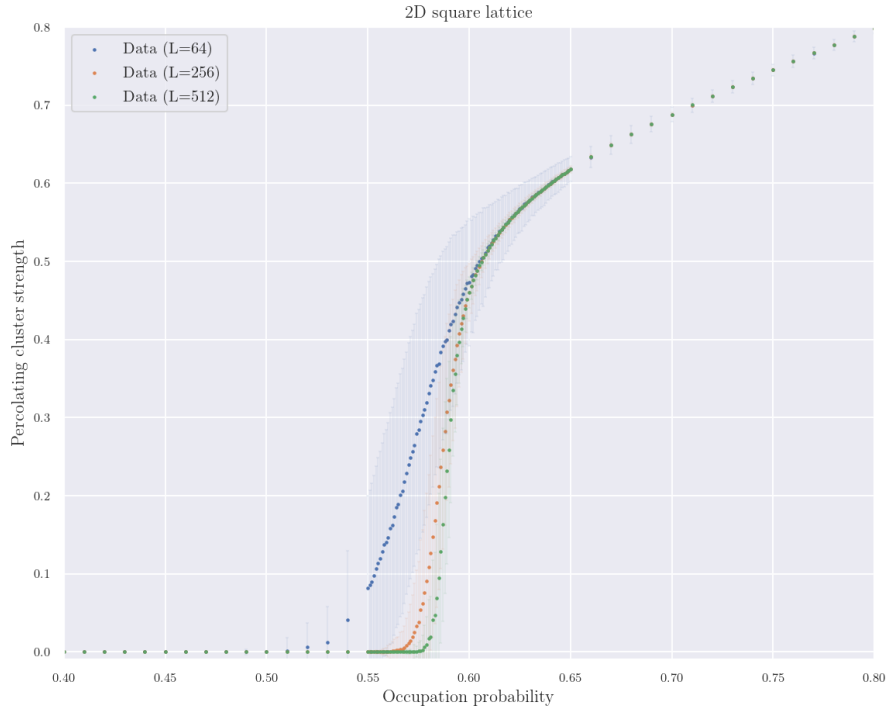


Figure 7: Percolating cluster strength versus occupation probability for various lattice sizes

2.4 MEAN CLUSTER SIZE

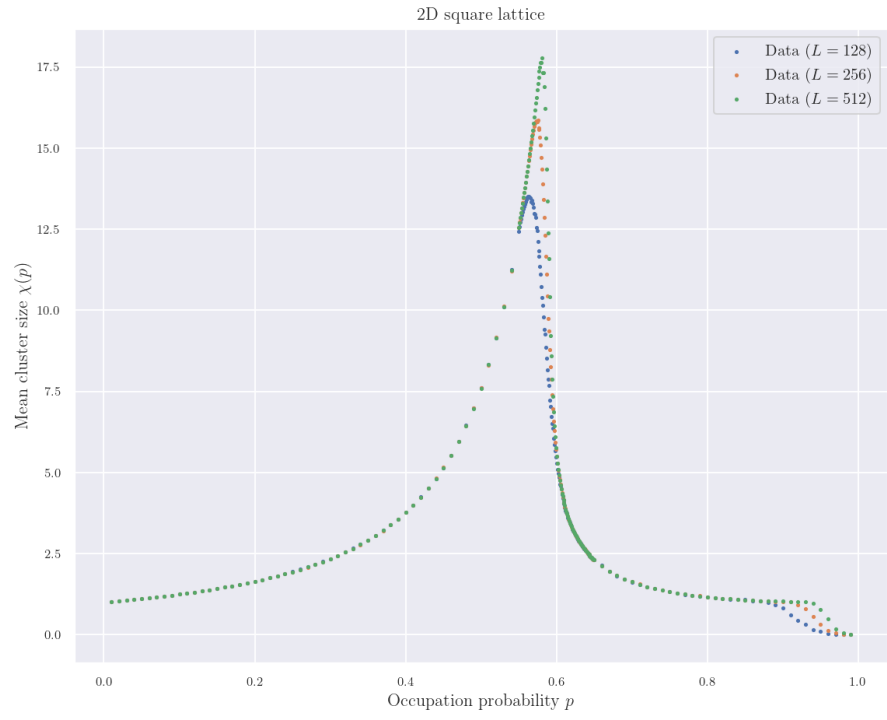


Figure 8: Mean finite cluster size versus occupation probability for various lattice sizes

2.5 CORRELATION FUNCTION AND CORRELATION LENGTH

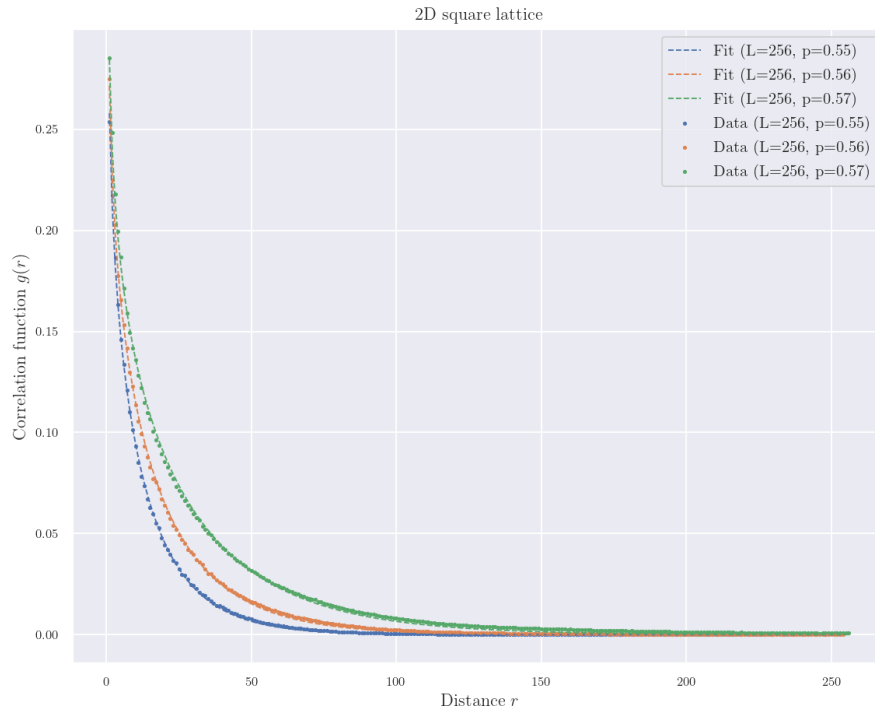


Figure 9: Correlation function for various sizes and occupation probabilities

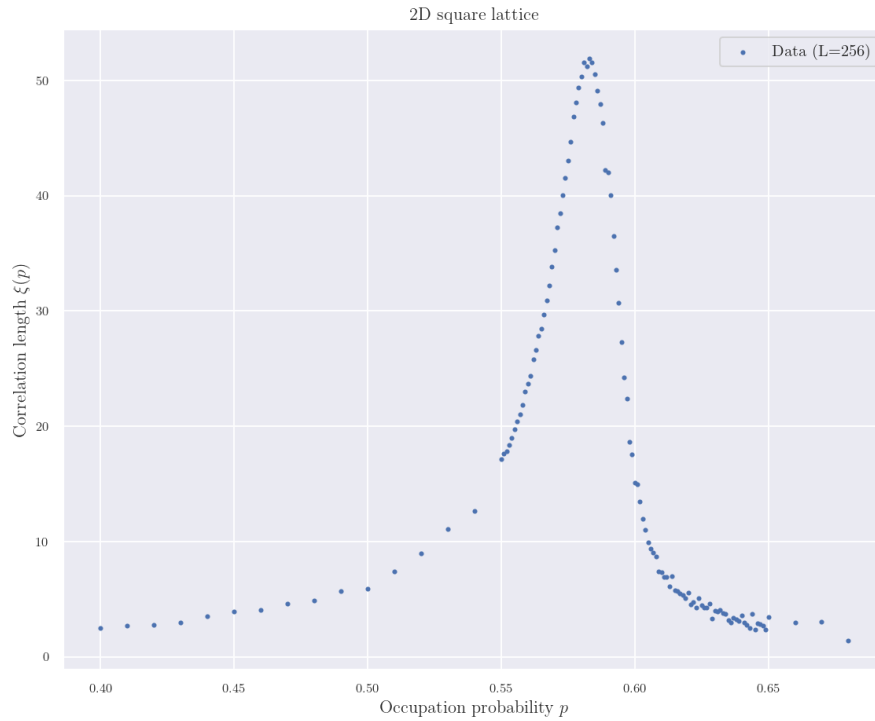


Figure 10: Correlation length for multiple lattice sizes

$$g(r) \propto r^{-\nu} e^{-\frac{r}{\xi}}$$

Part II

MANDELBROT PERCOLATION

Part III

APPENDIX



CODE

<https://github.com/alansammarone/musk>

DATA

https://github.com/alansammarone/MandelbrotPercolation/data/mysql_dump.tar.gz

TABLES

L	α	β
16	17.316	0.502909
24	22.9593	0.523217
32	28.7445	0.535866
48	37.8221	0.549603
64	46.2392	0.557582
96	61.2356	0.566495
128	75.4226	0.571311
192	102.451	0.576883
256	126.736	0.579948
294	139.832	0.581133

Table 1: Hyperbolic tangent least squares fit parameters

BIBLIOGRAPHY

- [1] Y. S. Cho and B. Kahng. "Discontinuous percolation transitions in real physical systems." In: *Phys. Rev. E* 84.050102 (2011).
- [2] P. Erdős and A. Rényi. "On random graphs, I." In: *Publ. Math. Debrecen* 6 (1959), pp. 290–297.
- [3] P. Erdős and A. Rényi. "On the evolution of random graphs." In: *Publ. Math. Inst. Hungar. Acad. Sci.* 5 (1960), pp. 17–61.
- [4] E. J. Friedman and A. S. Landsberg. "Construction and analysis of random networks with Explosive Percolation." In: *Phys. rev. Lett.* 103.255701 (2009).
- [5] F. Radicchi and S. Fortunato. "Explosive Percolation in scale-free networks." In: *Phys. Rev. Lett.* 103.168701 (2009).
- [6] O. Riordan and L. Warnke. "Achlioptas Process phase transitions are continuous." In: *The Annals of Applied Probability* 22 (2012), pp. 1450–1464.
- [7] M. Sahimi. *Applications of Percolation Theory*. 1st ed. Philadelphia, Pennsylvania: Taylor and Francis, 1994.
- [8] D. Stauffer and A. Aharony. *Introduction to Percolation Theory*. 2nd ed. Philadelphia, Pennsylvania: Taylor and Francis, 2003.
- [9] Yong Zhu and Xiaosong Chen. "Finite size scaling theory for percolation phase transition." In: *arXiv* 1710.02957 (2017).