In this paper we use visibility graphs as a tool to analyse the results of kinetic Monte Carlo (kMC) simulations of submonolayer deposition. We introduce an efficient algorithm for the computation of the visibility graph resulting from a kMC simulation and show that from the properties of the visibility graph one can determine the critical island size, establishing an effective tool for identifying the critical island size from experimental data, where the nucleation and growth rules must be inferred from images.

I. INTRODUCTION

Submonolayer deposition (SD) is a term used to describe the initial stages of thin film growth, such as under molecular beam epitaxy, where monomers are deposited onto a surface, diffuse and form large-scale structures (islands). Of particular interest is the mechanism for island nucleation, and how it is reflected in the statistical properties of the growing structures. SD is widely studied using kinetic Monte Carlo (kMC) simulations and it is recognised that under suitable conditions (described below), the statistical properties of the growing structures display scale invariance with distributions reflecting the underlying nucleation mechanism [1].

Below we will be considering point islands, i.e. islands whose extent and internal structure have been neglected. Point island approximations are often used in SD models, as they approximate SD accurately when the islands are ‘well separated’ [2]. We consider a one-dimensional point island model, where monomers are randomly deposited onto an initially empty lattice \( L \) at a deposition rate of \( F \) particles per unit time \( t \). The monomers diffuse at a rate \( D \) and islands nucleate when \( i + 1 \) monomers coincide at a lattice site. We call \( i \) the critical island size. We assume no monomers can evaporate from the lattice and so the coverage \( \theta \) can be defined as \( \theta = 100Ft/L\% \). \( \theta \) is chosen large enough for us to be in the aggregation regime (where scale-invariance is found) i.e. where the monomers are much more likely to aggregate into islands than nucleate into new islands. The appropriate value of \( \theta \) where the aggregation regime starts is dependent on \( i \) and the ratio \( R = D/F \).

All previous work on SD models has focused on the spatial distribution of islands or on the size statistics of their islands, see [2–4] for details and the scaled gap and island size distributions for different critical island sizes \( i \) illustrated in Figure 1 (for limited data). We define the gap between islands as the distance between two nucleated sites and the island size as the number of monomers at a particular site on the lattice (where the size of the island must be at least \( i+1 \)). It is worth noting that spatial distributions and size statistics of islands are negligibly affected by variations in the number of lattice sites \( L \), by variations in \( R \), or by the particular \( \theta \) in the aggregation regime [4].

One would like to combine the information contained in the spatial distribution of islands and in their size-statistics. A suitable tool, which could allow higher-dimensional extensions which we discuss in the conclusion section, is offered by visibility graphs introduced by Lacasa et al. [5] originally to bring the tools of network theory to bear on time-series analysis. Subsequently, visibility graphs have been used to analyse several rate series [6] and to make solar cycle predictions [7]. Using visibility graphs in the context of kMC simulations of SD allows us to employ complex network theory to SD. Here we will apply this approach, develop an efficient algorithm for processing the data, and show how characteristic metrics of the visibility graph can be used to identify the critical island size from the island and gap size data.

Briefly, in a visibility graph we connect each point \( P \)
with coordinates (location, size) (the top of our grey bars in Figure 2) to all other points that ‘are visible’ from \( P \) and analyse the resulting graph [5].

![Figure 2: An example of a visibility graph where the blue lines represent the edges in the network.](image)

II. THE VISIBILITY GRAPH ALGORITHM

First, we would like to describe an efficient algorithm for the computation of a visibility graph given \( n \) points in the plane, \( S = \{P_1, P_2, \ldots, P_n\} \), where \( P_j = (x_j, y_j) \), \( j \in \{1, \ldots, n\} \). The simplest way to construct the visibility graph is by considering points \( P_a, P_b \in S \) (assuming without loss of generality that \( x_a < x_b \)); then \( P_a \) and \( P_b \) are visible from each other if all points \( P_c \in S \) such that \( x_a < x_c < x_b \), satisfy

\[
y_c < y_b + \frac{y_b - y_a}{x_b - x_a}(x_c - x_b).
\]

To construct the visibility graph we need to consider all two-point subsets of \( S \), which gives us an algorithm with time complexity of \( C = \frac{1}{2}n^3 + O(n^2) \). As our kMC simulations produce up to \( 10^5 \) nucleated sites per simulation, this algorithm is impractical as one visibility graph takes nearly two hours to produce on a single core desktop PC. Hence, we aim to find an algorithm that is faster than the naïve one.

We collect the results needed for the construction of such an algorithm in the following claims. Throughout, we let \( P_a, P_b, P_c \in S \) be such that \( x_a < x_b < x_c \).

**Claim II.1** Let \( A = (a_{jk}) \) where \( j, k \leq n \) be adjacency matrix of the visibility graph. Then \( a_{jj} = 0, a_{jj+1} = 1 \) when \( j < n \), \( a_{jk} = a_{kj} \).

**Claim II.2** Let \( P_a \) and \( P_b \) be connected and \( y_a < y_b \). Then all points \( P_c \) such that \( y_c < y_b \) are not visible from \( P_a \).

**Claim II.3** Let \( P_a \) be connected to \( P_b \) and \( P_b \) be connected to \( P_c \). Then the slopes of the line segments connecting \( P_a \) to \( P_b \) and \( P_b \) to \( P_c \) are given by

\[
m_1 = \frac{y_b - y_a}{x_b - x_a} \quad \text{and} \quad m_2 = \frac{y_c - y_b}{x_c - x_b}, \quad \text{respectively}.
\]

Thus

1. if \( m_2 > m_1 \), \( P_c \) is visible from \( P_a \),
2. if \( m_2 \leq m_1 \), \( P_c \) is not visible from \( P_a \).

For two vectors \( \mu = (\mu_1, \mu_2) \) and \( \nu = (\nu_1, \nu_2) \) we define \( \mu \land \nu = \mu_1 \nu_2 - \mu_2 \nu_1 \).

**Claim II.4** Let \( P_a \) be connected to \( P_b \) and define \( \tilde{P}_c := (x_c, 0) \). Then \( P_c \) is visible from \( P_a \) if and only if

\[
t_1 = \frac{\gamma_2 \land \gamma_1}{\gamma_2 \cdot \gamma_3} \in [0, \infty) \quad \text{and} \quad t_2 = \frac{\gamma_1 \cdot \gamma_3}{\gamma_2 \cdot \gamma_3} \in [0, 1],
\]

where \( \gamma_1 = P_a - P_c, \gamma_2 = \tilde{P}_c - P_a, \) and \( \gamma_3 = -(P_b(2) - P_a(2), P_b(1) - P_a(1)) \), see Figure 3.

![Figure 3: An illustration of (II.4)](image)

A. New Visibility Graph Algorithm

We can now use Claims (II.1)-(II.4) to construct our visibility graph algorithm. We consider an arbitrary point \( P_j \in S \) and the vector of elements to the right of the main diagonal in the \( j \)-th row of the adjacency matrix of the visibility graph \( [a_{j,j+1}, a_{j,j+2}, \ldots, a_{j,n}] \).

Letting \( 2 \leq k \leq n - j \) we have:

- \( a_{j,j+i} = 1 \), by Claim (II.1).
- \( a_{j,j+k-1} = 1 \) and \( y_{j+i} < y_{j+k-1} \) where \( k \leq i \leq n - j \), then \( a_{j,j+i} = 0 \) by Claim (II.2).
- \( a_{j,j+k-1} = 1 \) then \( a_{j,j+k} = 1 \) if \( m_2 > m_1 \) and \( a_{j,j+k} = 0 \) otherwise by Claim (II.3).
- \( a_{j,j+k-1} = 0 \), then \( a_{j,j+k} = 1 \) if \( t_1 \in [0, \infty) \) and \( t_2 \in [0, 1] \) and \( a_{j,j+k} = 0 \) otherwise by Claim (II.4).

We continue this process for all \( P_j \in S \) and then use the final property from Claim (II.1) to complete our adjacency matrix.

Our new algorithm is around 15 times faster than the original and typically reduces computation time from two hours to eight minutes for the construction of a visibility graph from one kMC run.
III. CHARACTERISING THE VISIBILITY GRAPH

Thus, we start with a kMC simulation of SD in one-space dimension. Once the simulation is complete, we mark the location and the size (‘height’) of each nucleated island and construct the resulting visibility graph.

Our simulations were performed on lattices with \( L = 10^6 \) sites, \( R = 10^6 \) up to coverage of \( \theta = 200\% \) for different critical island sizes \( i \). (For \( i = 0 \) we set the spontaneous nucleation probability i.e. the chance a monomers becomes fixed to the lattice to \( p = 10^{-6} \).) We choose these conditions to guarantee we are in the aggregation regime and throughout the remainder of this paper we refer to these conditions as our ‘standard conditions’.

There are many ways to characterise a graph; these include criteria based on vertex degree, spectrum of the adjacency and other matrices defined from the graph, communicability and centrality indices [8]. Below we only analyse the vertex degree distribution and spectral gap in the adjacency matrix as these are sufficient to differentiate between visibility graphs corresponding to different critical island sizes \( i \).

A. Degree Distribution

We begin our characterisation of the visibility graph by considering the vertex degree distribution. Let \( n \) be the number of nodes in our visibility graph and \( m(k) \) be the number of nodes in our visibility graph with \( k \) connectivity; for simplicity we define \( q(k) := m(k)/n \). The vertex degree distributions of visibility graphs generated from kMC simulations (under our standard conditions) averaged over 50 simulations are shown in Figure 4.

![Figure 4: The vertex degree distributions of visibility graphs generated from kMC simulations on lattices with \( L = 10^6 \) sites, with \( R = 10^6 \) and up to coverage of \( \theta = 200\% \) when \( i = 0, 1, 2 \) and 3 and in the \( i = 0 \) case we let \( p = 10^{-6} \), averaging results over 50 runs.](image)

From Figure 4, we see that graphs corresponding to different values of \( i \) differ in the statistics of nodes having degree \( k \), particularly when \( 3 \leq k \leq 8 \). To investigate this finding further, we consider this specific region as shown in Figure 5. To emphasise the differences we connect the points with straight lines.

![Figure 5: The vertex degree distributions of visibility graphs generated from kMC simulations on lattices with \( L = 10^6 \) sites, with \( R = 10^6 \) and up to coverage of \( \theta = 200\% \) when \( i = 0, 1, 2 \) and 3 and in the \( i = 0 \) case we let \( p = 10^{-6} \), averaging results over 50 runs when \( 3 \leq k \leq 8 \).](image)

As expected, for every \( i \) considered, the degree distributions are exponential (see [9] for further details), however there are noticeable differences particularly for \( q(3) \) for different \( i \). Changes in \( R \) (when \( R = 10^7, 10^8 \) and \( 10^9 \)), \( L \) (when \( L = 10^7 \) and \( 10^8 \)) and \( \theta \) (when \( \theta = 100\% \)) have a negligible effect on the degree distributions confirming that we are operating in the aggregation (scaling) regime. This is consistent with the work on gap size, island size and spatial distributions, see [2].

To test if our values for \( q(3) \) (from Figure 5) can accurately predict \( i \), we generate a visibility graph from a kMC simulation, where \( i \) is assumed to be unknown, and compute the corresponding \( q(3) \) to predict \( i \). Our simulations were performed using a range of values for \( L, R \) and \( \theta \) as described above (these conditions guarantee we are in the aggregation regime; we will refer to them as ‘operational conditions’) and \( i \) was randomly chosen between 0 and 3, averaging results over 50 runs. We performed this process 50 times. We found that \( i \) was correctly predicted in 92% of cases. In addition, in all cases the predicted \( i \) was within 1 of the true value of \( i \).

B. Spectrum of the adjacency matrix

Next we consider the adjacency matrix of the visibility graph. We consider the first five eigenvalues of the adjacency matrix of our visibility graphs generated from kMC simulations under our standard conditions. As with the vertex degree distribution, we find consistent behaviour for each \( i \). We average the eigenvalues over 50 runs, as shown in Figure 6.

![Figure 6:](image)
and in the coverage of $\theta$ averaging results over 50 runs. Then to predict from a kMC simulation under our operational conditions nucleation mechanisms, we generate a visibility graph simulations when $p$ = $10^{-6}$, averaging results over 50 runs for the adjacency matrix.

As the $i = 0$ case is practically indistinguishable from the $i = 1$ case, to separate the two we consider the gap between the largest eigenvalue and the second largest eigenvalue from the adjacency matrix i.e. the spectral gap, which has been shown to be related to the connectivity of the graph [8]; these results are shown in Figure 7.

![Figure 6](image1.png)

**FIG. 6:** The eigenvalues of visibility graphs generated from kMC simulations on lattices with $L = 10^6$ sites, with $R = 10^6$ and up to coverage of $\theta = 200\%$ when $i = 0, 1, 2$ and 3 and in the $i = 0$ case we let $p = 10^{-6}$, averaging results over 50 runs for the adjacency matrix.

As the $i = 0$ case is practically indistinguishable from the $i = 1$ case, to separate the two we consider the gap between the largest eigenvalue and the second largest eigenvalue from the adjacency matrix i.e. the spectral gap, which has been shown to be related to the connectivity of the graph [8]; these results are shown in Figure 7.

Once again, we find that changes in $R$, $L$ and $\theta$ have a negligible effect on the eigenvalues and the gap between the eigenvalues.

As in the case of using $q(3)$ to distinguish between nucleation mechanisms, we generate a visibility graph from a kMC simulation under our operational conditions averaging results over 50 runs. Then to predict $i$ we use the largest eigenvalue to separate the case of $i = 2, 3$ from the rest, and finally, the spectral gap is used to differentiate the case of $i = 0$ from that of $i = 1$. We performed this process 50 times. We found that $i$ was correctly predicted in all cases.

**IV. CONCLUSIONS**

We have shown that the analysis of the properties of the visibility graph generated by a kMC simulation allows us to predict the underlying nucleation mechanism; as expressed in the critical island size $i$ both the degree distribution ($q(3)$) and the spectral gap in the adjacency matrix allow us reliably to predict the $i$ of the simulation. We have also created an efficient algorithm for processing the kMC position/size data. Therefore, we have created an effective characterisation process that can be applied to experimental data for SD in one dimension, such as island nucleation and growth on a stepped substrate [10]. This approach allows the molecular scale rules for nucleation and growth to be decided where they might otherwise be impossible to deduce. Our methods demonstrate that the visibility graph approach can usefully combine spatial and size data in a physically meaningful way, relating SD to network theory and opening up new approaches to SD description and classification.

An important question is how to extend this methodology to two and three-space dimensions. There are many possibilities to investigate one possible construction in two-space dimensions is as follows. Having performed a kMC simulation or obtained an experimental image of islands grown on the substrate, we inscribe all the islands in the unique circle $C$ of minimal radius and then from each island drop perpendiculars to a line $\Gamma$ tangent to the circle. This clearly induces a visibility graph on $\Gamma$. Work is underway to analyse the properties of the resulting visibility graphs dependence on the orientation $\Gamma$ and, whether as in the one-dimensional case, the graphs allow us to differentiate between the different nucleation mechanisms. Alternatively, we could consider the ideas developed in [11] that construct a visibility graph in 2 space dimensions among polyhedral obstacles. To consider the mechanisms developed in [11] we can not consider our islands to be point islands, however this approach may be useful in the context of extended islands.

---

[1] J. G. Amar and F. Family, Phys. Rev. Lett. 74, 2066 (1995).
[2] P. A. Mulheran, K. P. O’Neill, M. Grinfeld and W. Lamb, Phys. Rev. E 86, 051606 (2012).
[3] P. A. Mulheran and D. A. Robbie, Europhys. Lett. 49, 617-623 (2000).
[4] J. Blackman, P. A. Mulheran, Phys. Rev. B 54, 11681 (1996).
[5] L. Lacasa, B. Luque, F. Ballesteros, J. Luque and J. C. Nuño, Proc. Nat. Acad. Sci. 105 4972-4975 (2008).
[6] Y. Yang, J. Wang, H. Yang and J. Mang, Phys. Rev. A 388 4431-4437 (2009).
[7] Y. Zou, M. Small, Z. H. Liu and J. Kurths, Phys. New. J 16 013051 (2014).
[8] L. Lovász, OUP, Oxford 2007.
[9] D. Allen, Ph.D. thesis, University of Strathclyde, Glasgow, UK, thesis in preparation (2018).
[10] C. D. Pownall and P. A. Mulheran Phys. Rev. B 60 9037 (1999).
[11] T. Lozano-Pérez and M. A. Wesley, Commun. ACM 22
560-570 (1979).