

The Complexity of Flood Filling Games

Raphaël Clifford, Markus Jalsenius, Ashley Montanaro, and Benjamin Sach

Department of Computer Science, University of Bristol, UK
{clifford,markus,montanar,ben}@cs.bris.ac.uk

Abstract. We study the complexity of the popular one player combinatorial game known as Flood-It. In this game the player is given an $n \times n$ board of tiles where each tile is allocated one of c colours. The goal is to make the colours of all tiles equal via the shortest possible sequence of flooding operations. In the standard version, a flooding operation consists of the player choosing a colour k , which then changes the colour of all the tiles in the monochromatic region connected to the top left tile to k . After this operation has been performed, neighbouring regions which are already of the chosen colour k will then also become connected, thereby extending the monochromatic region of the board. We show that finding an optimal solution for Flood-It is **NP**-hard for $c \geq 3$ and that this even holds when the player can perform flooding operations from any position on the board. However, we show that this ‘free’ variant is in **P** for $c = 2$. We also prove that for an unbounded number of colours, Flood-It remains **NP**-hard for boards of height at least 3, but is in **P** for boards of height 2. Next we show how a $(c-1)$ approximation and a randomised $2c/3$ approximation algorithm can be derived, and that no polynomial time constant factor, independent of c , approximation algorithm exists unless **P=NP**. We then investigate how many moves are required for the ‘most demanding’ boards and show that the number grows as fast as $\Theta(\sqrt{cn})$. Finally, we consider boards where the colours of the tiles are chosen at random and show that for $c \geq 2$, the number of moves required to flood the whole board is $\Omega(n)$ with high probability.

1 Introduction

In the popular one player combinatorial game known as Flood-It, each tile of an $n \times n$ board is allocated one of c colours, where c is a parameter of the game. Two left/right/up/down adjacent tiles are said to be connected if they have the same colour and a (connected) region of the board is defined to be any maximal connected component. The standard version of the game starts with the player ‘flooding’ the region that contains the top left tile. The flooding operation simply involves changing the colour of all the tiles in the region to be some new colour. However, this also has the effect of connecting the newly flooded region to all neighbouring regions of this colour. The overall aim is to flood the entire board, that is connect all regions, in as few flooding operations as possible. Every flooding operation changes the colour of the region that contains the top left tile. Figure 1 gives an example of the first few moves of a game. The border shows the outline of the region which has so far been flooded.

In this paper, we investigate a number of questions inspired by Flood-It. We first show that not only are natural greedy approaches to the game bad, but in fact finding an optimal solution for Flood-It is **NP**-hard for $c \geq 3$, and that this also holds for a variant of the game we call Free-Flood-It where the player can perform flooding operations at any position on the board. On the other hand, we

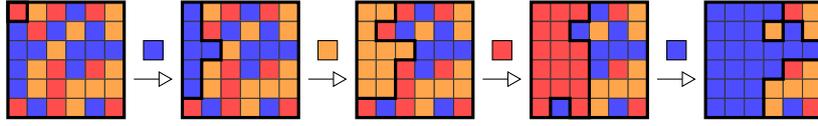


Fig. 1: A sequence of four moves on a 6×6 Flood-It board with 3 colours.

show that solving Free-Flood-It with $c = 2$ is in \mathbf{P} . We also consider the effect of changing the shape of the board, and prove that Flood-It remains \mathbf{NP} -hard for rectangular boards of height at least 3, with an unbounded number of colours, but is in \mathbf{P} for boards of height 2. Next we show how a $(c - 1)$ approximation and a randomised $2c/3$ approximation algorithm for Flood-It can be derived. However, no polynomial time constant factor, independent of c , approximation algorithm exists unless $\mathbf{P}=\mathbf{NP}$. We then consider how many moves are required for the most demanding boards and show that the number grows as fast as $\Theta(\sqrt{cn})$. Finally, we investigate boards where the colours of the tiles are chosen at random and give a simple proof that for $c \geq 3$, the number of moves required to flood the whole board is $\Omega(n)$ with high probability. We then observe that the same result can in fact be proven for $c \geq 2$ by appealing to previous deep results in percolation theory [2,5].

History and related work: Perhaps the most famous recent hardness result involving a popular game is the \mathbf{NP} -completeness of Tetris [3]. Flood-It seems to be a somewhat newer game than Tetris, first making its appearance online in early 2006 courtesy of a company called Lab Pixies. Since then numerous versions have become available for almost every conceivable platform. We have very recently become aware of a sketch proof by Elad Verbin posted on a blog of the \mathbf{NP} -hardness of Flood-It with 6 colours [14]. Although our work was completed independently, it is interesting to note that there is some similarity to the techniques used in our \mathbf{NP} -hardness proof for $c \geq 3$ colours.

Independently of this work, Fleischer and Woeginger have studied a closely related game to Flood-It, known as Honey-Bee [4]. This game is also based around repeatedly applying a flood filling operation on a grid. The main differences are that the grid is hexagonal and may contain barriers, and also that there is a two-player variant of the game. In this variant, two players start flood filling from opposite corners, and the goal is to control more of the board than your opponent. Fleischer and Woeginger focus on the computational complexity of Honey-Bee, and consider a number of generalisations of the single player game to different classes of graphs. They prove that some generalisations are \mathbf{NP} -hard, while others are in \mathbf{P} . Again, there is some similarity in the techniques used in one of their \mathbf{NP} -hardness proofs, although we note that this proof does not immediately apply to Flood-It without some modification. Fleischer and Woeginger also show that the two-player game on arbitrary graphs is \mathbf{PSPACE} -complete.

Another related game whose computational complexity has been studied in detail is known as Clickomania [1]. A rectangular board is initialised in the same way as in Flood-It. The move permitted is for the player to remove a

chosen connected monochromatic component of at least two tiles after which any blocks above it will fall down as far as they can. Finding an optimal solution to Clickomania has been shown to be **NP**-hard for two or more columns and five or more colours, or five or more columns and three or more colours.

Flood-It can also be thought of as a model for a number of different (possibly not entirely) real world applications. For example, our results supplement that of recent work on zombie infestation [11] if one regards the flooding operation as one where the minds of neighbouring non-zombies are infected by those who have already been turned into zombies. A separate but no less significant line of research considers the complexity of tools commonly provided with Microsoft Windows. Previous work has shown that aspects of Excel [6] and even Minesweeper [8] are **NP**-complete. Our work extends this line of research by showing that flood filling in Microsoft’s Paint application is also **NP**-hard.

1.1 Notation and definitions

Let $B_{n,c}$ be the set of all $n \times n$ boards with at most c colours. We write $m(B)$ for the minimum number of moves required to flood a board $B \in B_{n,c}$. We will refer to rows and columns in a board in the usual manner. We further denote the colour of the tile in row i and column j as $B[i, j]$; colours are represented by integers between 1 and c . Throughout we assume that $2 \leq c \leq n^2$.

We define a *diamond* to be a diamond-shaped subset of the board (see Figure 2a). These structures are used throughout the paper. The centre of the diamond is a single tile and the *radius* is the number of tiles from its centre to its leftmost tile. We write r -diamond to denote a diamond of radius r . A single tile is therefore a 1-diamond. For $i \in \{1, \dots, r\}$, the i th *layer* of an r -diamond is the set of tiles at board distance $i - 1$ from its centre. We will also consider diamonds which are cropped by intersection with the board edges as in Figure 2b.

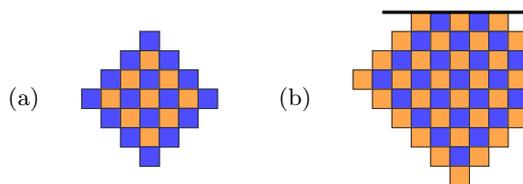


Fig. 2: (a) An alternating 4-diamond and (b) a cropped 6-diamond.

2 A greedy approach is bad

An obvious strategy for playing the Flood-It game is the greedy approach. There are two natural greedy algorithms: (1) we pick the colour that results in the largest gain (number of acquired tiles), or (2) we choose the colour dominating the perimeter of the currently flooded region. It turns out that both these approaches can be surprisingly bad.

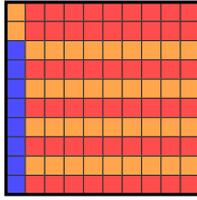


Fig. 3: A 10×10 board where a greedy approach is bad.

To see this, let B be the 10×10 board on three colours illustrated in Figure 3. The number of moves required to flood B is three. However, either greedy approach given would first pick the colours appearing on the horizontal lines before finally choosing to flood the left-hand vertical column. In both cases, this requires 10 moves to fill the board. It should be clear how this example can easily be extended to arbitrarily large $n \times n$ boards. In general, the greedy algorithm will make n moves, while the optimal algorithm will still make only three.

3 The complexity of Flood-It

Let c -FLOOD-IT denote the problem which takes as input an $n \times n$ board B of c colours and outputs the minimum number of moves $m(B)$ in a Flood-It game that are required to flood B . Similarly, let c -FREE-FLOOD-IT denote the generalised version of c -FLOOD-IT in which we are free to flood fill from an arbitrary tile in each move. Although we have seen that a straightforward greedy algorithm fails, it is not too far-fetched to think that a dynamic programming approach would solve these problems efficiently, but the longer one ponders over it, the more inconceivable it seems. To aid frustrated Flood-It enthusiasts, we prove in this section that both c -FLOOD-IT and c -FREE-FLOOD-IT are indeed **NP**-hard, even when the number of colours is as small as three. Interestingly, we will see that 2-FREE-FLOOD-IT is in **P**.

To show **NP**-hardness, we reduce from the *shortest common supersequence* problem, denoted SCS, which is defined as follows. The input is a set S of k strings over an alphabet Σ . A *common supersequence* s of the strings in S is a string such that every string in S is a subsequence of s . The output is the length of a shortest common supersequence of the strings in S . The decision version of SCS takes an additional integer ℓ and outputs yes if the shortest common supersequence has length at most ℓ , otherwise it outputs no.

Maier [10] showed in 1978 that the decision version of SCS is **NP**-complete if the alphabet size $|\Sigma| \geq 5$. A couple of years later, R aih a and Ukkonen [12] extended this result to hold for $|\Sigma| \geq 2$. For a long time, various groups of people tried to approximate SCS but no polynomial-time algorithm with guaranteed approximation bound was to be found. It was not until 1995 that Jiang and Li [7] settled this open problem by proving that no polynomial-time algorithm can achieve a constant approximation ratio for SCS, unless **P** = **NP**. The following lemma proves the **NP**-hardness of both c -FLOOD-IT and c -FREE-FLOOD-IT when the number of colours is at least four. The inapproximability of both prob-

lems also follows immediately from the approximation preserving nature of the reduction. We will need a more specialised reduction for the case $c = 3$, which is given in Lemma 2.

Lemma 1 *For $c \geq 4$, c -FLOOD-IT and c -FREE-FLOOD-IT are **NP**-hard (and the decision versions are **NP**-complete). Further, for an unbounded number of colours c , there is no polynomial-time constant factor approximation algorithm, unless $\mathbf{P} = \mathbf{NP}$.*

Proof. The proof is split into two parts; first we prove the lemma for c -FLOOD-IT in which we flood fill from the top left tile in each move, and in the second part we generalise the proof to c -FREE-FLOOD-IT in which we can flood fill from any tile in each move.

We reduce from an instance of SCS that contains k strings s_1, \dots, s_k each of length at most w over the alphabet Σ . Suppose that $\Sigma = \{a_1, \dots, a_r\}$ contains $r \geq 2$ letters and let $\Sigma' = \{b_1, \dots, b_r\}$ be an alphabet with r new letters. For $i \in \{1, \dots, k\}$, let s'_i be the string obtained from s_i by inserting the character b_j after each a_j and inserting the character b_1 at the very front. For example, from the string $a_3a_1a_4a_3$ we get $b_1a_3b_3a_1b_1a_4b_4a_3b_3$.

Let $\Sigma \cup \Sigma'$ represent the set of $2r$ colours that we will use to construct a board B . First, for $i \in \{1, \dots, k\}$, we define the $|s'_i|$ -diamond D_i such that the j th layer will contain only one colour which will be the j th character from the right-hand end of s'_i . Thus, the colour of the outermost layer of D_i is the first character of s'_i (which is b_1 for all strings) and the centre of D_i is the last character of s'_i . The reason why we intersperse the strings with letters from the auxiliary alphabet Σ' is to ensure that no two adjacent layers of a diamond have the same colour. This property is crucial in our proof. Let B be a sufficiently large $n \times n$ board constructed by first colouring the whole board with the colour b_1 and then placing the k diamonds D_i on B such that no two diamonds overlap. Since each of the k diamonds has a radius of at most $2w + 1$, we can be assured that n never has to be greater than $k(4w + 1)$.

Suppose that s is a shortest common supersequence of s_1, \dots, s_k and suppose its length is ℓ . We will now argue that the minimum number of moves to flood B is exactly 2ℓ , first showing that 2ℓ moves are sufficient. Let s' be the 2ℓ -long string obtained from s by inserting the character b_j after each a_j . We make 2ℓ moves by choosing the colours in the same order as they appear in s' . Note that we flood fill from the top left tile in each move. From the construction of the diamonds D_i it follows that all diamonds, and hence the whole board, are flooded after the last character of s' has been processed.

It remains to be shown that at least 2ℓ moves are necessary to flood B . Let s'' be a string over the alphabet $\Sigma \cup \Sigma'$ that specifies a shortest sequence of moves that would flood the whole board B . From the construction of the diamonds D_i it follows that the string obtained from s'' by removing every character in Σ' is a common supersequence of s_1, \dots, s_k and therefore has length at least ℓ . By symmetry (replace every a_j with b_j in the strings s_1, \dots, s_k), the string obtained from s'' by removing every character in Σ has length at least ℓ as well. Thus, the length of s'' is at least 2ℓ .

Since the decision version of SCS is **NP**-complete even for a binary alphabet Σ , it follows that c -FLOOD-IT is **NP**-hard for $c \geq 4$, and the decision version is **NP**-complete. The inapproximability result in the statement of the lemma follows immediately from the reduction.

Now we show how to extend these results to c -FREE-FLOOD-IT. The reduction from SCS is similar to the previously presented reduction. However, instead of constructing only one board B , we construct $2kw + 1$ copies of B and put them together to one large $n' \times n'$ board B' . If necessary in order to make B' a square, we add sufficiently many $n \times n$ boards that are filled only with the colour b_1 . Note that $(2kw + 1)n$ and hence $(2kw + 1)k(4w + 1)$ is a generous upper bound on n' .

From the construction of B' it follows that exactly 2ℓ moves are required to flood B' if we flood fill from the top left tile in each move; all copies of B will be flooded simultaneously. The question is whether we can do better by flood filling from tiles other than the top left one (or any tile in its connected component). That is, can we do better by picking a tile inside one of the diamonds? We will argue that the answer is no. First note that $2\ell \leq 2kw$. Suppose that we do flood fill from a tile inside some diamond D for some move. This move will clearly not affect any of the other diamonds on B' . Suppose that this move would miraculously flood the whole of D in one go so that we can disregard it in the subsequent moves. However, there were originally $2kw + 1$ copies of D , which is one more than the absolute maximum number of moves required to flood B' , hence we can use a recursive argument to conclude that flood filling from a tile inside a diamond will do us no good and would only result in more moves than if we choose to flood fill from the top left tile in each move. \square

The reduction in the previous proof is approximation preserving, which allowed us to prove that there is no efficient constant factor approximation algorithm. We reduced from an instance of SCS by doubling the alphabet size, resulting in instances of c -FLOOD-IT and c -FREE-FLOOD-IT with $c \geq 4$ colours. To establish **NP**-hardness for $c = 3$ colours, we need to consider a different reduction. We do this in the lemma below by reducing from the decision version of SCS over a binary alphabet to the decision versions of 3-FLOOD-IT and 3-FREE-FLOOD-IT. Note that this reduction is not approximation preserving.

Lemma 2 *3-FLOOD-IT and 3-FREE-FLOOD-IT are **NP**-hard (and the decision versions are **NP**-complete).*

Proof. We reduce from an instance of the decision version of SCS on k strings s_1, \dots, s_k of length at most w over the binary alphabet $\{1, 2\}$ and an integer ℓ . The yes/no question is whether there exists a common supersequence of length at most ℓ .

For $i \in \{1, \dots, k\}$, let s'_i be the string obtained from s_i by inserting the new character 3 at the front of s_i and after each character of s_i . Let the set $\{1, 2, 3\}$ represent the colours that we will use to construct a board B . First, for each of the k strings s'_i we define the diamond D_i exactly as in the proof of Lemma 1 (see Figure 4a). We define R to be the following rectangular area of the board

of width $4\ell + 5$ and height $2\ell + 3$. Let x be the middle tile at the bottom of R . Around x we have layers of concentric half rectangles (see Figure 4b). We refer to these layers as *arches*, with the first arch being x itself. As demonstrated in the figure, the first arch has the colour 1 and the second arch has the colour 2. All the remaining odd arches have the colour 3, and all the remaining even arches are coloured 2 everywhere except for the tile above x which has the colour 1. As described in detail below, the purpose of these arches is to control which minimal sequences of moves would flood B .

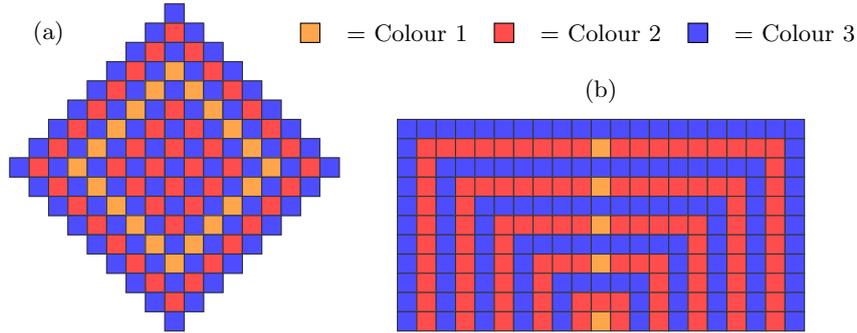


Fig. 4: An example of (a) a diamond and (b) a rectangle constructed in the proof of Lemma 2.

Let B be a sufficiently large $n \times n$ board constructed as follows. First colour the whole board with the colour 3. Then, at the bottom of B starting from the left, place $2\ell + 3$ copies of R one after another without any overlaps. Finally place the k diamonds D_i on B such that no two diamonds overlap and no diamond overlaps any copy of R . Figure 5 illustrates a board B with $\ell = 2$ and $k = 10$. Since a diamond has a radius of at most $2w + 1$ and $\ell \leq kw$, $k(4w + 1) + (2kw + 3)(4kw + 5)$ is an upper bound on n .

The reason why we place copies of R on the board B is to make sure that at least $2\ell + 2$ moves are required to flood B , even in the absence of diamonds. To see this, suppose first that we flood fill from the top left square in each move. From the definition of the arches of R , disregarding the diamonds on B , a minimal sequence of moves will consist of ℓ 1s or 2s interspersed with a total of $\ell - 1$ 3s, followed by the three moves 3, 2 and 1, respectively. Note that only one copy of R on B would be enough to achieve this. However, having several copies of R on B does not affect the minimum number of moves as all copies will get flooded simultaneously. The idea with the $2\ell + 3$ copies of R is to make sure that at least $2\ell + 2$ moves are required to flood B even when we are allowed to choose which tile to flood fill from in each move. To see this, suppose that we choose to flood fill from a tile inside one of the copies of R . Since there are $2\ell + 3$ copies, similar reasoning to the end of the proof of Lemma 1 tells us that we will do worse than $2\ell + 2$ moves.

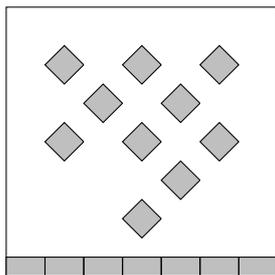


Fig. 5: A board constructed in the proof of Lemma 2.

We will now argue that the number of moves required to flood B is $2\ell + 2$ if and only if there is a common supersequence of s_1, \dots, s_k of length at most ℓ . We choose to flood fill from the top left tile in each move.

Suppose first that there is a common supersequence s of length $\ell' \leq \ell$. Let s' be the string s followed by $\ell - \ell'$ 1s. Let s'' be the $(2\ell + 2)$ -long string obtained from s' by inserting a 3 after each character of s' and adding the two additional characters 2 and 1 to the end. We make $2\ell + 2$ moves by choosing the colours in the same order as they appear in s'' . Note that all diamonds are flooded after $2\ell'$ moves, and by the last move we have also flooded every copy of R , and hence the whole board B .

Suppose second that B can be flooded in $2\ell + 2$ moves. The centre of each diamond has the colour 3 and therefore the first 2ℓ moves flood the diamonds. The subsequence of these first 2ℓ moves induced by the colours 1 and 2 is an ℓ -long common supersequence of s_1, \dots, s_k . \square

We can now summarise Lemmas 1 and 2 in the following theorem.

Theorem 3. *For $c \geq 3$, c -FLOOD-IT and c -FREE-FLOOD-IT are **NP**-hard (and the decision versions are **NP**-complete). Further, for an unbounded number of colours c , there is no polynomial-time constant factor approximation algorithm, unless $\mathbf{P} = \mathbf{NP}$.*

For two colours, 2-FLOOD-IT is trivially in \mathbf{P} , but it is not that obvious what the complexity of 2-FREE-FLOOD-IT is. The next theorem settles this question, by showing that an optimal strategy for any instance of 2-FREE-FLOOD-IT consists of flooding from the same tile in each move.

Theorem 4. *2-FREE-FLOOD-IT is in \mathbf{P} .*

Proof. We first consider the case where we are allowed to flood fill from exactly two distinct tiles of the board. At the end of the proof we consider the case where flooding from any tile is allowed.

Suppose there exists a shortest sequence of moves S that floods the board from only two tiles t_1 and t_2 . Suppose also that t_1 and t_2 belong to different connected components during the first $m_1 + m_2$ moves but become connected in the $(m_1 + m_2 + 1)$ th move, where m_1 is the number of flood filling operations from t_1 and m_2 is the number of flood filling operations from t_2 . Suppose without

loss of generality that in move $m_1 + m_2 + 1$, we flood fill from t_1 . Let t'_1 and t'_2 be two adjacent tiles such that after $m_1 + m_2$ moves, t_1 and t'_1 belong to the same monochromatic region, and t_2 and t'_2 belong to the same monochromatic region. Let P_1 be a simple path from t_1 to t'_1 in the board with the monochromatic connected components $\alpha_0, \dots, \alpha_{m_1}$, such that the i th flood filling move from t_1 merges α_i with the monochromatic region that contains t_1 . Thus, $t_1 \in \alpha_0$ and the whole path P_1 is monochromatic after m_1 flood filling operations from t_1 . We define a path P_2 from t_2 to t'_2 similarly. Let $\beta_0, \dots, \beta_{m_2}$ be the monochromatic connected components of P_2 . Figure 6 illustrates the two paths P_1 and P_2 .

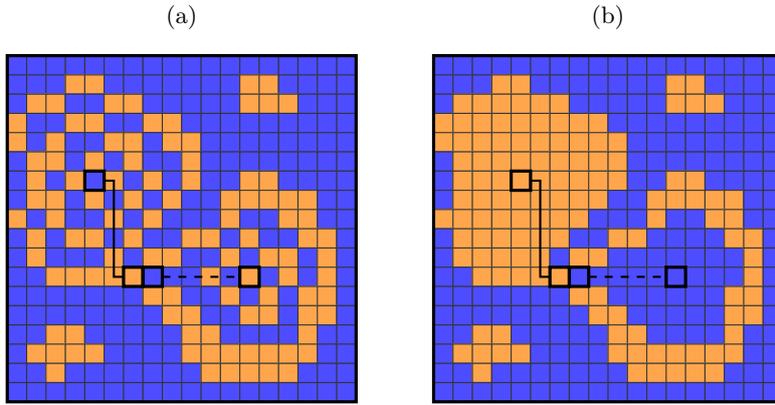


Fig. 6: A board (a) before and (b) after $m_1 + m_2$ moves as discussed in the proof of Theorem 4. The solid and dashed paths give P_1 and P_2 respectively. In left-to-right order, the emphasised tiles are t_2, t'_2, t'_1 and t_1 .

We now show that the area flooded after the first $m_1 + m_2 + 1$ moves of S can be flooded with $m_1 + m_2 + 1$ flood filling moves from one single tile t_3 . Let P_3 be the path P_1 concatenated with a reversed copy of P_2 . Thus, the monochromatic connected components γ_i of P_3 are $\gamma_0 = \alpha_0, \gamma_1 = \alpha_1, \dots, \gamma_{m_1} = \alpha_{m_1}, \gamma_{m_1+1} = \beta_{m_2}, \gamma_{m_1+2} = \beta_{m_2-1}, \dots, \gamma_{m_1+m_2+1} = \beta_0$. Let t_3 be a tile in γ_{m_2} and consider a series of flood filling moves from this tile: after the first m_2 moves, t_1 and t_3 are connected, and after the first $m_1 + 1$ moves, t_2 and t_3 are connected. Once a tile t is in the same monochromatic component as t_3 , flooding from t_3 is equivalent to flooding from t . Thus, after a total of $m_1 + m_2 + 1$ flood filling moves from t_3 , we have effectively performed $m_1 + 1$ flood filling moves from t_1 and m_2 flood filling moves from t_2 . This is exactly what the first $m_1 + m_2 + 1$ moves of S do. Hence we can replace the moves in S by flooding from a single tile t_3 .

Finally we deal with the case where we are allowed to flood fill from any tile. Consider a shortest sequence S of moves that flood the board and suppose that we flood fill from the tiles t_1, \dots, t_r , for $r > 2$. Suppose without loss of generality that the first merge of any of these tiles is when we flood fill from t_1 , which connects t_1 with $t_2, \dots, t_{r'}$, where $2 \leq r' \leq r$. Let m_i be the number of flood filling operations that have taken place from t_i before this merge. The following sequence S' of moves will flood the board in at most $|S|$ moves but flood fills

from only $r - 1$ tiles. For $i = 3, \dots, r$, first perform m_i flood filling operations from t_i . Instead of flooding from t_1 and t_2 separately, we use the result above and flood fill from a different tile t' . Thus, by the next $m_1 + m_2 + 1$ moves we have connected $t_1, \dots, t_{r'}$. The subsequent moves of S' follow those of S , where any move at t_1 or t_2 is replaced with a move at t' . Inductively we reduce the number of tiles to flood fill from to a single tile. The conclusion is that we can solve 2-FREE-FLOOD-IT by attempting to flood the entire board from each tile of the board in turn, which requires only polynomial time. \square

4 The complexity of constant height boards

So far we have analysed the complexity of FLOOD-IT on square shaped $n \times n$ boards. A natural question to ask is: what is the complexity of c -FLOOD-IT on an $h \times n$ board, where the height h is a fixed constant? We denote this problem by (c, h) -FLOOD-IT and the ‘free’ variant by (c, h) -FREE-FLOOD-IT, analogously.

$(c, 1)$ -FLOOD-IT is trivially in \mathbf{P} , and Fleischer and Woeginger have shown (personal communication) that $(c, 1)$ -FREE-FLOOD-IT is also in \mathbf{P} . We will show that of $(c, 2)$ -FLOOD-IT on a $2 \times n$ board remains in \mathbf{P} . However, the complexity of $(c, 2)$ -FREE-FLOOD-IT remains unresolved. Before stating this result we will prove in Theorem 6 that when the number of colours is unbounded and $h \geq 3$ then both (c, h) -FLOOD-IT and (c, h) -FREE-FLOOD-IT are \mathbf{NP} -hard.

For the c -FLOOD-IT problem on a square $n \times n$ board with $c \geq 4$ we gave a reduction from the shortest common supersequence problem (SCS) which embedded a number of diamond structures into a board filled with a single background colour. Each diamond represented one of the strings in the SCS instance. The problem with this reduction on an $h \times n$ board is that a string of length ℓ was represented by a diamond with height $4\ell - 1$. This is not possible if $h < 4\ell - 1$. However, Timkovskii proved [13] that the SCS problem remains \mathbf{NP} -hard even when the length of the strings is constrained to be at most 2, and the alphabet size is unbounded. Inspection of the proof of Lemma 1 shows that (c, h) -FLOOD-IT is \mathbf{NP} -hard (and the decision version \mathbf{NP} -complete) when $h \geq 8$ and the number of colours is unbounded. Naively, it would appear that $h = 7$ suffices in the proof of Lemma 1 as it allows enough height to embed a diamond representing a string of length 2 as is required. However, for the reduction to be valid we also need to leave at least one row of space above the diamonds so that all diamonds can be flooded simultaneously on any move.

To reduce the board height required for our \mathbf{NP} -hardness proof further we reduce the height of the diamond structures used in the reduction. Recall that the reduction begins by doubling the length of all strings in a way that ensures that no string contains a character which is followed immediately by another occurrence of the same character. We now show in Lemma 5 that the SCS problem remains \mathbf{NP} -hard even when all the strings are of the form ab where $a, b \in \Sigma$ and $a \neq b$. The proof is by reduction from the SCS problem with the constraint that all strings have length at most 2. This result allows us to remove the doubling step and reduce the height of the diamond structures, resulting in Theorem 6 which gives the desired result.

Lemma 5 *The SCS problem is NP-hard when all the strings are of the form ab where $a, b \in \Sigma$ and $a \neq b$.*

Proof. Let S be an instance of SCS that contains k strings s_1, \dots, s_k each of length $w \leq 2$ over the alphabet Σ . We abuse notation by referring to S as both the instance and the set of k strings. We begin by assuming that S contains a string of length 1. Without loss of generality let s_k be such a string. Let S' be the instance of SCS formed by the $k - 1$ strings s_1, \dots, s_{k-1} . There are two cases to consider. In the first case, the single character, a , in s_k occurs in some string s_1, \dots, s_{k-1} . Therefore any common supersequence of S' contains an a and hence is also a common supersequence of S . Further, as S is a superset of S' , any common supersequence of S is a common supersequence of S' . Hence $|\text{SCS}(S)| = |\text{SCS}(S')|$. In the second case, the single character, a , in s_k does not occur in s_1, \dots, s_{k-1} . A common supersequence for S can therefore be found by inserting a at the end of the shortest common supersequence for S' . Hence $|\text{SCS}(S)| \leq |\text{SCS}(S')| + 1$. Further, any common supersequence for S must contain an a and is also a common supersequence for S' . Therefore by removing the a we have that $|\text{SCS}(S)| \geq |\text{SCS}(S')| + 1$. Hence in this case $|\text{SCS}(S)| = |\text{SCS}(S')| + 1$. Repeated application of the above technique gives a poly-time reduction from the SCS problem with strings of length $w \leq 2$ to the SCS problem with strings of length $w = 2$. Therefore we have that the latter is also NP-hard.

We now redefine S to be an instance of SCS that contains k strings s_1, \dots, s_k each of length exactly 2 over the alphabet Σ . We begin by assuming that S contains a string of the form aa where $a \in \Sigma$. Without loss of generality let s_k be such a string. Let S' be the instance of SCS formed by the $k - 1$ strings s_1, \dots, s_{k-1} and new strings $s_{k+1} = aa'$ and $s_{k+2} = a'a$ where a' does not occur in S . First consider the shortest common supersequence of S , which must contain aa as a subsequence. By inserting a' between these two occurrences of a , we obtain a common supersequence of S' of length $|\text{SCS}(S)| + 1$. Therefore $|\text{SCS}(S')| \leq |\text{SCS}(S)| + 1$. Now consider the shortest common supersequence of S' , which must contain either $aa'a$ or $a'aa'$ as a subsequence. In the former case by removing the a' symbol we obtain a common supersequence of S and have that $|\text{SCS}(S')| \geq |\text{SCS}(S)| + 1$. In the latter, when we remove the two occurrences of a' we obtain a common supersequence of s_1, \dots, s_{k-1} of length $|\text{SCS}(S')| - 2$. This sequence contains exactly one occurrence of a , and by inserting a second we obtain a common supersequence of S of length $|\text{SCS}(S')| - 1$. Therefore $|\text{SCS}(S')| = |\text{SCS}(S)| + 1$. Repeated application of the above technique gives a poly-time reduction from the SCS problem with strings of length 2 to the SCS problem with strings of the form ab where $a, b \in \Sigma$ and $a \neq b$. Therefore we have that the latter is also NP-hard. \square

Theorem 6. *(c, h) -FLOOD-IT and (c, h) -FREE-FLOOD-IT are NP-hard when $h \geq 3$ and the number of colours c is unbounded (and the decision versions are NP-complete).*

Proof. First observe that the decision versions of both problems are in NP because the unconstrained versions, c -FLOOD-IT and c -FREE-FLOOD-IT, are in

NP. We begin by considering the (c, h) -FLOOD-IT problem for $h \geq 3$. We reduce from an instance of SCS on k strings s_1, \dots, s_k over the alphabet Σ and an integer ℓ . The strings are constrained to have the form ab where $a, b \in \Sigma$ and $a \neq b$. Let B be a $h \times n$ board filled with a single background colour where $n = 4k + 1$. For each symbol in Σ we have a corresponding distinct colour in addition to the background colour. For each string $s_i = a_i b_i$ we embed a ‘half’ diamond against the bottom edge of the board. The half diamond consists of a single tile of colour b_i (the inner layer), surrounded on all three sides by a tile of colour a_i (the outer layer). This is illustrated in Figure 7 for $h = 3$.

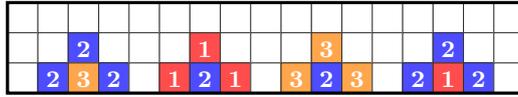


Fig. 7: An example of a board constructed in the proof of Theorem 6. In left-to-right order, the strings embedded are “23”, “12”, “32” and “21”. The shortest common supersequence is “2132”.

Observe that as $h \geq 3$ and $n = 4k + 1$, all the half diamonds can be placed so that the outer layer of each half diamond is surrounded by the background colour. Therefore on any move, the outer layer of any half diamond can be flooded. Further observe that for all i , as $a_i \neq b_i$ the diamond for s_i is flooded if and only if the move sequence contains $a_i b_i$ as a subsequence. Therefore a move sequence floods the board if and only if it is a common supersequence of s_1, \dots, s_k , so $|\text{SCS}(S)|$ equals the length of the shortest move sequence which floods the board. As this reduction can be implemented in polynomial time, we have that (c, h) -FLOOD-IT problem is **NP**-hard with an unbounded number of colours.

We now consider the (c, h) -FREE-FLOOD-IT problem for $h \geq 3$. **NP**-hardness follows by the same argument as for the **NP**-hardness of c -FREE-FLOOD-IT for $c \geq 4$ given in the proof of Lemma 1. We increase the size of the board (horizontally) and embed $2k + 1$ copies of each half diamond. We observe that any flood-filling move begun from a tile in an unflooded half diamond floods only tiles in that half diamond. This ensures that any move sequence which floods the board and contains moves begun from tiles in an unflooded half diamond either contains at least $2k + 1$ moves or contains redundant moves. In either case, it is not minimal. \square

We finally show that $(c, 2)$ -FLOOD-IT is in **P**.

Theorem 7. *For any $c \geq 1$, c -FLOOD-IT on a $2 \times n$ board is in **P**. More precisely, the running time is $O(n)$.*

Proof. Suppose that B is a $2 \times n$ board and c is the number of colours. We say that a tile t on B is *marked* if it has colour c_t and no other tile in the columns strictly to the right of t has the colour c_t . A column is marked if it contains a marked tile.

The key observation, which holds on a $2 \times n$ board, is that if the marked tiles are flooded then so is the whole board B . To see this, note that when a marked tile t of colour c_t is flooded, all other tiles of the colour c_t that have not yet been flooded are to the left of t and therefore adjacent to the flooded region. Hence they will be flooded when t is flooded. Thus, we ask for the shortest sequence of moves that would flood the marked tiles.

A *shortest path* to a tile t denotes a shortest sequence of flood filling operations that includes t in the flooded region. If t is already included in the flooded region, then the length of the shortest path to t is 0.

One might think that a solution to c -FLOOD-IT on a $2 \times n$ board would be to go from one marked tile to the next in left-to-right order using shortest paths. Although this is correct, we must be a little careful with which shortest paths we choose. The following procedure floods the marked tiles in the smallest number of moves possible.

Beginning of procedure. Let i be the leftmost marked column such that i contains a marked tile t that has not yet been flooded. Let t' be the other tile in column i . We have two cases.

Case 1 (t' is unmarked). Let m and m' be the lengths of the shortest paths to t and t' , respectively. Note that $|m - m'| \leq 1$. We consider two subcases.

Case 1a ($m \leq m'$). Flood using the sequence of colours found along the shortest path to t , then go to the beginning of the procedure. Correctness: Flooding t' before t means that we are bound to flood t at a later stage. Once t' is flooded we can never do worse by flooding t immediately. Thus, flooding t' before t and then flooding t takes a total of at least $m + 1$ moves. However, flooding t takes m moves and we are not necessarily forced to spend an extra move on flooding t' , which is not a marked tile.

Case 1b ($m > m'$). Flood using the sequence of colours found along the shortest path to t' and then flood t . Then go to the beginning of the process. Correctness: Flooding t takes at least $m' + 1$ steps, even if we do not go via t' . Since all remaining marked tiles are to the right of column i , we should therefore flood t' before t . Once t' is flooded, we can never do worse by flooding t immediately.

Case 2 (t' is marked). Flood using the sequence of colours found along the shortest of the shortest paths to t or t' . Then flood the remaining tile in column i . Then go to the beginning of the process. Correctness: Both t and t' must eventually be flooded. Once one of them is flooded, there is no reason to wait to flood the other.

Using for example dynamic programming, the shortest path to a tile t on a $2 \times n$ board can be computed in time linear in the distance between the flooded region and t . We note that the shortest paths are always calculated between the rightmost end of the flooded region and a marked column i . Since the flooded region is always extended to column i in each step of the procedure, the total

running time of computing the shortest paths is linear in n . Hence the running time of the whole process is $O(n)$. \square

5 Approximating the number of moves

As we have seen, c -FLOOD-IT and c -FREE-FLOOD-IT are not efficiently approximable to within a constant factor for an unbounded number of colours c . However, a $(c - 1)$ -approximation for c -FLOOD-IT, $c \geq 3$, can easily be obtained as follows. Suppose that B is a board on the colours $1, \dots, c$. Clearly, if we repeatedly cycle through the sequence of colours $1, \dots, c$ then B will be flooded after at most $c \times m(B)$ moves. We can do a little better by first cycling through the ordered sequence of colours $1, \dots, c$ and then repeatedly alternating between a cycle of the sequence $(c - 1), \dots, 1$ and a cycle of $2, \dots, c$ until there are only two distinct colours left on the board, after which we alternate between the two remaining colours. Note that there are always exactly two distinct colours left before the final move. The board B is guaranteed to be flooded after at most $c + (c - 1)(m(B) - 2) + 1 \leq (c - 1)m(B)$ moves, which gives us a $(c - 1)$ -approximation algorithm.

A randomised approach with an expected number of moves of approximately $2c/3 \times m(B)$ is obtained as follows. Suppose that s is a minimal sequence of colours that floods B (flood filling from the top left square in each move). We shuffle the c colours and process them one by one. If B is not flooded then we shuffle again and repeat. Thus, if $m(B) = 1$ then the algorithm takes c moves. If $m(B) = 2$ then it takes $c + \frac{1}{2}c = 3c/2$ expected number of moves; with probability a half, both moves in s appear (in correct order) during the first c moves, otherwise we need c additional moves for the last move in s . We generalise this as follows. Let $T(m)$ be (an upper bound on) the expected number of moves it takes to produce a fixed sequence of m moves. We have $T(m) = c + \frac{1}{2}T(m - 1) + \frac{1}{2}T(m - 2)$. Solving the recurrence with the values of $T(1)$ and $T(2)$ above gives us a solution in which $T(m)$ is asymptotically $(2c/3)m$ for a fixed c .

6 General bounds on the number of moves

Recall that we denote the minimum number of moves which flood some board B as $m(B)$. In this section we investigate bounds on the maximum $m(B)$ over all boards in $B_{n,c}$ which we denote $\max\{m(B) \mid B \in B_{n,c}\}$. Intuitively, this can be seen as the minimum number of moves to flood the ‘worst’ board in $B_{n,c}$.

For motivation, consider an $n \times n$ checker board of two colours as shown in Figure 8. First observe that as the board has only two colours, the player has no choice in their next move. Consider a diagonal of tiles in the direction top-right to bottom-left where the 0th diagonal is the top-left corner. Further observe that move k floods exactly the k th diagonal, so the total number of moves is $2(n - 1)$. Thus we have shown that $\max\{m(B) \mid B \in B_{n,c}\} \geq 2(n - 1)$.

We now give an overview of a simple algorithm which floods any board in $B_{n,c}$ in at most $c(n - 1)$ moves. The algorithm performs n stages. The purpose

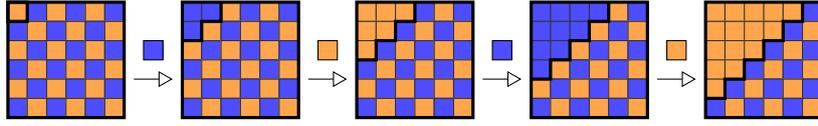


Fig. 8: Progression of a 6×6 checker board.

of the i th stage is to flood the i th row. Stage i repeatedly picks the colour of the leftmost tile in row i which is not in the flooded region, until row i is flooded.

First observe that Stage 1 performs at most $n - 1$ moves to flood row i (we can flood at least one tile of row 1 per move). When the algorithm begins Stage $i \geq 2$, observe that row $i - 1$ is entirely flooded as well as any tiles in row i which match the colour of row $i - 1$. Therefore when a new colour is selected, all tiles in row i of this colour become flooded. Hence at most $c - 1$ moves are performed by Stage i . Summing over all rows, this gives the desired bound that $\max\{m(B) \mid B \in B_{n,c}\} \leq c(n - 1)$. Observe that from the previous example with the checker board on $c = 2$ colours, the bound $c(n - 1)$ is tight. Thus, the checker board is the ‘worst’ board in $B_{n,2}$.

As motivation, we have given weak bounds on $\max\{m(B) \mid B \in B_{n,c}\}$. We now tighten these bounds for large c by providing a better algorithm for flooding an arbitrary board. We will also give a description of ‘bad’ boards which require many moves to be flooded. It will turn out that $\max\{m(B) \mid B \in B_{n,c}\}$ is asymptotically $\Theta(\sqrt{cn})$ for increasing n and c .

Theorem 8. *There exists a polynomial time algorithm for FLOOD-IT which can flood any $n \times n$ board with c colours in at most $2n + (\sqrt{2c})n + c$ moves.*

Proof. For a given integer ℓ (to be determined later), we partition the board horizontally into $\ell + 1$ contiguous sections, denoted S_0, \dots, S_ℓ from top to bottom, as follows. Let $q = \lfloor n/\ell \rfloor$ and $r = n \bmod \ell$. Section S_0 consists of the first $\lceil q/2 \rceil$ rows, S_1, \dots, S_r contain $(q + 1)$ rows each (if $r > 0$), and $S_{r+1}, \dots, S_{\ell-1}$ contain q rows each (if $r < \ell - 1$). Section S_ℓ contains $\lfloor q/2 \rfloor$ rows. See Figure 9 for an illustration. We let $y(i)$ denote the final row of S_i .

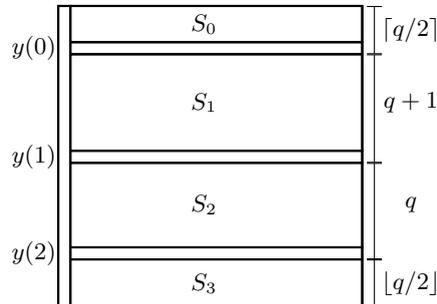


Fig. 9: The board decomposition used in the proof of Theorem 8.

The algorithm performs the following three stages.

- Stage 1. Flood the first column.
- Stage 2. Flood row $y(x)$ for all $0 \leq x < \ell$.
- Stage 3. Cycle through the c colours until the board is flooded.

The correctness of our algorithm is immediate as Stage 3 ensures that the board is flooded by cycling colours. Stage 1 can be implemented to perform at most $n - 1$ moves as argued for the simple algorithm above. Similarly, Stage 2 can be completed in $\ell(n - 1)$ moves. We now analyse Stage 3.

First consider S_0 . At the start of Stage 3, row $y(0)$ is entirely in the top-left region, so a single cycle of the c colours suffices to expand the region to include row $y(0) - 1$. Each subsequent cycle of c colours expands the region to include an additional row. Therefore, after $c(\lceil q/2 \rceil - 1) \leq cq/2$ moves of Stage 3, all rows above $y(0)$ are included in the top left region. Similarly, the section S_ℓ will be included in the top-left region as it contains $\lfloor q/2 \rfloor \leq q/2$ rows.

Now consider section S_i for some $0 < i < \ell$. Observe that there are at most q rows in S_i which are not already completely in the top-left section (after stage 2). Further observe that any cycle of c colours expands the region to include *two* more of these rows. One row is gained from the region bordering the top of the section (which is in the top-left region from stage 2). The second is gained from the region bordering the top of the section (which is also in the top-left region from stage 2). Therefore after at most $c\lceil q/2 \rceil$ moves of Stage 3 the board is flooded.

Over all three stages this gives a total of at most $n + \ell n + c\lceil q/2 \rceil$ moves. We pick $\ell = \lceil \sqrt{c/2} \rceil$ to minimise this number of moves. By recalling that $q = \lfloor n/\ell \rfloor$ and simplifying we have that this total is less than $2n + \sqrt{2c}n + c$ moves as required. \square

Theorem 9. *For $2 \leq c \leq n^2$, there exists an $n \times n$ board with (up to) c colours which requires at least $\sqrt{c-1}n/2 - c/2$ moves to flood.*

Proof. Suppose first that c is even. For a given integer $r \geq 1$, let $D_{(x,y)}$ be an r -diamond where odd layers are coloured x and even layers are coloured y . Any board containing $D_{(x,y)}$ requires at least r moves of colours x and y . Further, observe that as long as the centre of $D_{(x,y)}$ is in the board, even if it is cropped by at most two edges of the board, at least r moves of colours x and y are still required (see Figure 2b). We refer to such an r -diamond as *good*. The central idea is to populate the board with good r -diamonds, $D_{(1,2)}, D_{(3,4)}, \dots, D_{(c-1,c)}$. As each r -diamond uses two colours (or one of the two colours if $r = 1$) which do not occur in any other diamond, the board must take at least $rc/2$ moves to flood.

It is not difficult to show that at least $(n^2 - r^2)/(2r^2)$ good r -diamonds can be embedded in an $n \times n$ board. An example of such a packing for a 20×20 board is given in Figure 10 (which shows only the edges of diamonds and not their colouring). This scheme generalises well to an $n \times n$ board but the details are omitted in the interest of brevity.

We now take $r = \lfloor n/\sqrt{c} \rfloor < n/2$ and note that $r \geq 1$. As $r < n/2$, the r -diamonds are cropped by at most two board edges as required. Therefore we

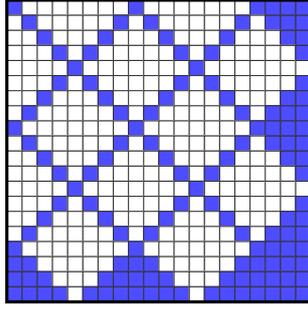


Fig. 10: 4-diamonds packed in a 20×20 board.

have at least $(n^2 - r^2)/(2r^2) \geq c/2 - 1/2$ good r -diamonds in our board. However, as the number of good r -diamonds is an integer, this is at least $c/2$ as required. Therefore, the number of moves required to flood this board is at least $rc/2 > n\sqrt{c}/2 - c/2$.

Finally, in the case that c is odd we proceed as above using $c - 1$ of the colours to give the stated result. \square

The next corollary is immediate from Theorems 8 and 9.

Corollary 10 $(\sqrt{c-1}n - c)/2 \leq \max\{m(B) \mid B \in B_{n,c}\} \leq 2n + \sqrt{2c}n + c$.

7 Random boards

In this section, we try to understand the complexity of a random Flood-It board – that is, a board where each tile is coloured uniformly at random. This question is of both theoretical and practical interest. A common initialisation for Flood-It is to pick the colours of tiles at random and the game designer will surely be keen to know if they are likely to have chosen an instance whose solution is trivially short. The option of having to solve every created instance to test for this possibility is also likely to be unattractive, especially given the complexity results shown in this paper. Intuitively, one would expect random boards to usually require a large number of moves to flood. Determining how many moves are actually needed turns out to be closely related to a body of research in *percolation theory*, the study of connected clusters in random graphs.

Indeed, a problem in percolation theory that is essentially equivalent to the question of the number of moves required for a random Flood-It board has been solved quite recently by Chayes and Winfield [2], and independently Fontes and Newman [5]. In our terminology, their result was that a random $n \times n$ Flood-It board with $c \geq 2$ colours requires $\Omega(n)$ moves with high probability. The proofs are lengthy and use some deep previous results in percolation theory.

We now present a greatly simplified proof of the results of [2,5], in the case that $c \geq 3$. Formally, our result is as follows.

Theorem 11. *Let $B \in B_{n,c}$ be a board where the colour of each tile is chosen uniformly at random from $\{1, \dots, c\}$. Then, for $c \geq 4$, $\Pr[m(B) \leq 2(3/10 - 1/c)(n - 1)] < e^{-\Omega(n)}$. For $c = 3$, $\Pr[m(B) \leq (n - 1)/22] < e^{-\Omega(n)}$.*

In order to prove this theorem, we will use two lemmas concerning paths in Flood-It boards. Let P be a simple path in a Flood-It board, i.e. a simple path on the underlying square lattice¹, where tiles are vertices on the path. Note that a path of length k includes $k + 1$ tiles. We say that a simple path P is *non-touching* if every tile in P is adjacent to at most two tiles that are also in P . Define the *cost* of P , $\text{cost}(P)$, to be the number of maximal monochromatic connected components of the path, minus one (so a monochromatic path has cost 0).

Lemma 12 *For any $B \in B_{n,c}$, there is a non-touching path from $(1, 1)$ to (n, n) with cost at most $m(B)$.*

Proof. For $m(B) = 0$ there is nothing to prove, so consider a strategy for completing B which uses $m(B) > 0$ moves. Label every tile $t \in B$ with an integer $m(t)$ between 0 and $m(B)$ that indicates the number of the move which changed the colour of t to be the colour of tile $(1, 1)$. Then, for each $i \geq 1$, there is a connected component labelled with i which has at least one neighbour labelled with $i - 1$. As the label of (n, n) is at most $m(B)$, and the label of $(1, 1)$ is 0, there is a simple path from $(1, 1)$ to (n, n) with cost at most $m(B)$. This path can be taken to be non-touching, because any pair of adjacent tiles (t_1, t_2) that are on the path but not connected by it correspond to a loop in the path that can be removed without increasing the cost. \square

Lemma 13 *For any integer $\ell \geq 3$, there are at most $4 \cdot 7^{(\ell-1)/2} < 2 \cdot (\sqrt{7})^\ell$ non-touching paths of length ℓ from any given tile.*

Proof. Let $T(\ell)$ denote the maximum number of non-touching paths of length ℓ starting from any given tile. $T(\ell)$ can be straightforwardly upper bounded by $4 \cdot 3^{\ell-1}$ for $\ell \geq 1$, as with each step of the path, aside from the first, there are at most 3 choices of direction. We get a tighter bound by analysing a few steps on a non-touching path P . Consider the i th vertex on P , for some $i \geq 2$. As P is simple, there are at most 3 choices for the $(i + 1)$ th vertex of the path. For vertex $i + 2$, if the previous two steps were in the same direction, there are at most 3 more choices. On the other hand, if the previous two were in different directions, there are only at most 2 choices (otherwise, the path would go back on itself, and would not be non-touching). In total, there are only at most 7 possible options for vertices $i + 1, i + 2$. Therefore, for any $\ell \geq 3$, we have $T(\ell) \leq 4 \cdot 7^{(\ell-1)/2}$. \square

The last result we will need is the following standard Chernoff bound.

Fact 14 *Let $X_i, 1 \leq i \leq m$, be independent 0/1-valued random variables with $\Pr[X_i = 1] = p$. Then $\Pr[\frac{1}{m} \sum_{i=1}^m X_i \geq p + \epsilon] \leq e^{-D(p+\epsilon||p)m} \leq e^{-2\epsilon^2 m}$, where $D(x||y)$ is the Kullback-Leibler divergence $D(x||y) = x \ln(x/y) + (1 - x) \ln((1 - x)/(1 - y))$.*

¹ Simple paths on square lattices have been intensively studied, and are known as *self-avoiding walks* [9]. There are known upper bounds, which are slightly stronger than Lemma 13, on the number of self-avoiding walks of a given length; however, we avoid these here to keep our presentation elementary.

We are finally ready to prove Theorem 11.

Proof (of Theorem 11). For any $k \geq 0$, and for any board B such that $m(B) \leq k$, by Lemma 12 there exists a non-touching path from $(1, 1)$ to (n, n) with cost at most k . So consider an arbitrary non-touching path P in B of length ℓ between these two tiles, and let P_i denote the i th tile on the path, for $1 \leq i \leq \ell + 1$. Note that $\ell \geq 2(n - 1)$. Then $\text{cost}(P) = |\{i : P_{i+1} \neq P_i\}|$, or equivalently $\text{cost}(P) = \ell - |\{i : P_{i+1} = P_i\}|$. Define the 0/1-valued random variable X_i by $X_i = 1 \Leftrightarrow P_{i+1} = P_i$. Then, as the colours of tiles are uniformly distributed, $\Pr[X_i = 1] = 1/c$ for all i , and

$$\Pr[\text{cost}(P) \leq k] = \Pr\left[\sum_{i=1}^{\ell} X_i \geq \ell - k\right] \leq e^{-D(1-k/\ell \| 1/c)\ell},$$

where we use Fact 14. Thus, using the union bound over all paths of length at least $2(n - 1)$ from $(1, 1)$ to (n, n) , we get that the probability that there exists *any* path of cost at most k is upper bounded by

$$2 \sum_{\ell=2(n-1)}^{\infty} (\sqrt{7})^{\ell} e^{-D(1-k/\ell \| 1/c)\ell} = 2 \sum_{\ell=2(n-1)}^{\infty} e^{((1/2) \ln 7 - D(1-k/\ell \| 1/c))\ell}, \quad (1)$$

where we use the estimate for the number of paths which was derived in Lemma 13. In the final part of the proof, we consider the cases $c \geq 4$ and $c = 3$ separately.

First suppose that $c \geq 4$. We take $k = 2(3/10 - 1/c)(n - 1) \leq (3/10 - 1/c)\ell$, as in the statement of the theorem, and use $D(1 - k/\ell \| 1/c) \geq 2(1 - k/\ell - 1/c)^2$ (from Fact 14) to obtain the bound

$$2 \sum_{\ell=2(n-1)}^{\infty} e^{((1/2) \ln 7 - 2(1-k/\ell-1/c)^2)\ell} \leq 2 \sum_{\ell=2(n-1)}^{\infty} e^{((1/2) \ln 7 - 49/50)\ell}.$$

As $49/50 > (1/2) \ln 7 \approx 0.973$, this sum is exponentially small in n .

Lastly, suppose that $c = 3$. In this case, our choice of k above is negative. Instead we take $k = (n - 1)/22$, which implies $1 - k/\ell \geq 43/44$. In order to obtain a sufficiently tight bound on $D(1 - k/\ell \| 1/c)$, we use the explicit formula in Fact 14 to show that $D(43/44 \| 1/3) > 0.974 > (1/2) \ln 7$, which implies that there is a bound in Equation (1) which is exponentially small in n . This completes the proof. \square

8 Conclusion and open problems

We have shown that, for three or more colours, FLOOD-IT is **NP**-hard. However, for two colours, the relaxed version of the problem termed FREE-FLOOD-IT in which we are allowed to flood fill from any tile of the board remains in **P**. Some interesting open questions remain. First, the complexity of solving FREE-FLOOD-IT on a height 2 board remains unresolved. Second, we conjecture that the true lower bound for random boards is $\Omega(\sqrt{cn})$, matching the general upper bound. Interestingly, the percolation theory techniques that we are aware of do not appear to allow for super-linear lower bounds of the sort that would be required.

9 Acknowledgements

AM was funded by an EPSRC Postdoctoral Research Fellowship. MJ was supported by the EPSRC. We are grateful to Dave Arthur for producing an implementation of Flood-It, complete with examples of our NP-hardness reductions. We would also like to thank Leon Atkins, Aram Harrow, Tom Hinton and Alex Popa for many helpful and encouraging discussions.

References

1. T.C. Biedl, E.D. Demaine, M.L. Demaine, R. Fleischer, L. Jacobsen, and J.I. Munro. The complexity of Clickomania. In *More games of no chance*, volume 42 of *MSRI Publications*, pages 389–404. Cambridge University Press, 2002.
2. L. Chayes and C. Winfield. The density of interfaces: A new first-passage problem. *Journal of Applied Probability*, 30(4):851–862, 1993.
3. E.D. Demaine, S. Hohenberger, and D. Liben-Nowell. Tetris is hard, even to approximate. *Computing and Combinatorics*, pages 351–363, 2003.
4. R. Fleischer and G. Woeginger. An algorithmic analysis of the Honey-Bee game. In *Proc. Fun with Algorithms 2010*, pages 178–189, 2010.
5. L. Fontes and C. Newman. First passage percolation for random colorings of \mathbb{Z}^d . *The Annals of Applied Probability*, 3(3):746–762, 1993.
6. K. Iwama, E. Miyano, and H. Ono. Drawing Borders Efficiently. *Theory of Computing Systems*, 44(2):230–244, 2009.
7. T. Jiang and M. Li. On the approximation of shortest common supersequences and longest common subsequences. *SIAM Journal of Computing*, 24(5):1122–1139, 1995.
8. R. Kaye. Minesweeper is NP-complete. *The Mathematical Intelligencer*, 22(2):9–15, 2000.
9. N. Madras and G. Slade. *The Self-Avoiding Walk*. Birkhauser, 1996.
10. D. Maier. The complexity of some problems on subsequences and supersequences. *Journal of the ACM*, 25(2):322–336, 1978.
11. P. Munz, I. Hudea, J. Imad, and R.J. Smith. When zombies attack!: Mathematical modelling of an outbreak of zombie infection. In *Infectious Disease Modelling Research Progress*, pages 133–150. Nova Science, 2009.
12. K.-J. R  ih   and E. Ukkonen. The shortest common supersequence problem over binary alphabet is NP-complete. *Theoretical Computer Science*, 16:187–198, 1981.
13. V. G. Timkovskii. Complexity of common subsequence and supersequence problems and related problems. *Cybernetics and Systems Analysis*, 25(5):565–580, 1989.
14. Is this game NP-hard?, May 2009. <http://valis.cs.uiuc.edu/blog/?p=2005>.