PRINCIPLES OF COMPLEX SYSTEMS

HW07 WRITE-UP

David W. Landay
University of Vermont
Graduate Student, Comp. Systems and Data Sci.

0.1 Problem 1:

In class, we made our way through a discrete version of a toy HOT model of forest fires. Carlson and Doyle's 1999 paper "Highly optimized tolerance: A mechanism for power laws in design systems", revolves around the equivalent continuous model's derivation.

$$\text{From the paper we have that } p(x) \propto A^{-\gamma}(x) \tag{1}$$

$$\text{For convenience, we will simply write } A^{-\gamma}$$

Using the relationship from (**1**), we may demonstrate the results from the table in the Carlson and Doyle paper: Let $p(x) = c\,x^{-(q-1)}$. Then,

$$p(x) = c\,x^{-(q-1)}$$

$$\implies \left(\frac{A^{-\gamma}}{c}\right) = x^{-(q-1)}$$

$$\implies \left(\frac{A^{-\gamma}}{c}\right)^{-(q-1)} = x$$

$$\implies A^{\frac{\gamma}{q+1}} \propto x \tag{2}$$

From the relationship found in (**2**), we can solve for $P_{\geq}(x)$ and $P_{\geq}(A)$:

$$\text{Recall that } \int_{p^{-1}(A^{-\gamma})}^{\infty} p(x)\,dx = P_{\geq}(p^{-1}(A^{-\gamma}))$$

$$= \int_{p^{-1}(A^{-\gamma})}^{\infty} c\,x^{-(q-1)}\,dx \qquad \text{(by def.)}$$

$$= -\frac{c}{q}\left(x^{-q}\right)\Big|_{A^{\frac{\gamma}{q+1}}}^{\infty} \qquad \text{(1.a.i from 1)}$$

Figure 1: $P_{\geq}(x)$

As $x \to \infty$, then $P_{\geq}(x) = -\frac{c}{q}\left(x^{-q}\right) \to 0$. So assessing the definite integral, we get

$$\frac{c}{q}\left(0 + (A^{\frac{\gamma}{q+1}})^{-q}\right)$$

$$= \frac{c}{q} A^{-\frac{q\gamma}{q+1}} \therefore \tag{1.a.ii}$$

Let $p(x) = c\,e^{-x}$, then by (**1**) we may let $p(x) = A^{-\gamma} = c\,e^{-x}$

$$\implies \left(\frac{A^{-\gamma}}{c}\right) = e^{-x}$$

Taking the natural logarithm of both sides yields

$$\ln(A^{-\gamma}) \propto -x$$

$$\implies -\ln(A^{-\gamma}) \propto x$$

$$\implies \gamma \ln(A) \propto x \tag{3}$$

Similar to before, (**3**) allows us to solve for $\int_{p^{-1}(A^{-\gamma})}^{\infty} p(x)\,dx$ to find $P_{\geq}(x)$ and $P_{\geq}(A)$. We get

$$\int_{p^{-1}(A^{-\gamma})}^{\infty} c\,e^{-x}\,dx$$

$$= -c\,e^{-x}\Big|_{\gamma \ln(A)}^{\infty} \tag{1.b.i from 3}$$

Figure 2: $P_{\geq}(x)$

Notice that as $x \to \infty$, $e^{-x} \to 0$. Hence, the solution to the definite integral becomes

$$-c\, e^{-x} \Big|_{\gamma \ln(A)}^{\infty} = -c \left(0 - e^{-\gamma \ln(A)} \right)$$

$$\implies P_{\geq}(A) = c\, A^{-\gamma} \therefore \tag{1.b.ii}$$

Finally, let $p(x) = c\, e^{-x^2}$. Then by (**1**), $A^{-\gamma} = c\, e^{-x^2}$. Hence, we may say

$$A^{-\gamma} = c\, e^{-x^2}$$

$$\implies A^{-\gamma} \propto e^{-x^2}$$

Taking the natural logarithm of both sides, we may find x with respect to A:

$$\implies \ln\left(A^{-\gamma} \right) \propto -x^2$$

$$\implies -\gamma \ln(A) \propto -x^2$$

$$\implies \gamma \ln(A) \propto x^2$$

$$\implies \sqrt{\gamma \ln(A)} \propto x \therefore \tag{4}$$

Solving for $\int_{p^{-1}(A^{-\gamma})}^{\infty} p(x)\, dx$ to find $P_{\geq}(x)$ and $P_{\geq}(A)$, we notice that the particular integral in question is not solvable analytically. However, we can re-write the integral in the form $\int u\, dv$ to approximate through integration by parts. Recall that the integration by parts formula can be derived from the product rule:

Let f and g be two functions of x, then the product rule gives us

$$(f\, g)' = f'\, g + f\, g'$$

by integrating both sides, we obtain

$$\int_a^b (fg)'\,dx = \int_a^b f'g + fg'\,dx$$

$$\implies fg = \int_a^b f'g\,dx + \int_a^b fg'\,dx$$

$$\implies fg - \int_a^b f'g\,dx = \int_a^b fg'\,dx$$

For simplicity, we can make the following substitutions:

$$\begin{aligned}&\text{Let } u = f(x), v = g(x),\\ &du = f'(x), \text{ and } dv = g'(x)\end{aligned}$$

Then we have

$$\int_a^b u\,dv = uv - \int_a^b v\,du$$

We can now write the definite integral $\int_{p^{-1}(A^{-\gamma})}^{\infty} p(x)\,dx = c\int_{p^{-1}(A^{-\gamma})}^{\infty} e^{-x^2}\,dx$ by substituting $u = e^{-x^2}$, and $dv = dx$ to get

$$c\int_{p^{-1}(A^{-\gamma})}^{\infty} e^{-x^2}\,dx = c\int_{p^{-1}(A^{-\gamma})}^{\infty} u\,dv$$

This implies that $du = u' = -2x\,e^{-x^2}$ and $v = \int_{p^{-1}(A^{-\gamma})}^{\infty} dv = x$. For now we will ignore labelling the bounds and the constant c. Continuing with integration by parts, we get

$$\int_{p^{-1}(A^{-\gamma})}^{\infty} u\,dv$$

$$= x\,e^{-x^2} + 2\int x^2 e^{-x^2}\,dx \qquad \text{(int. by parts)}$$

Performing another substitution, we let $t = x^2 e^{-x^2}$, $ds = dx$, $dt = -2x^3 e^{-x^2}$, and $s = x$.

Thus, we get that

$$= x\, e^{-x^2} + 2\,(\,t\, s - \int s\, dt\,)$$

$$= x\, e^{-x^2} + 2\, x^3\, e^{-x^2} + r(\,x\,)$$

$$= e^{-x^2}\,(\,x + 2x^3\,) + r(\,x\,)$$

$$= \frac{1}{x}\, e^{-x^2}\,(\,x^2 + 2x^4\,) + r(\,x\,)$$

where r is a function of x, representing the remainder term of the infinite integral expansion. From here, we may conclude by saying that

$$\frac{1}{x}\, e^{-x^2} \propto \int u\, dv = \int_{p^{-1}(\,A^{-\gamma}\,)}^{\infty} e^{-x^2}\, dx$$

Replacing our bounds and constant, c, we get

$$\implies \frac{c}{x}\, e^{-x^2}\Big|_{\sqrt{\gamma \ln(\,A\,)}}^{\infty} \propto \int_{p^{-1}(\,A^{-\gamma}\,)}^{\infty} e^{-x^2}\, dx\, dv \qquad \text{(1.c.i from 4)}$$

Figure 3: $P_{\geq}(x)$

Notice that as $x \to \infty$, then $\frac{c}{x}\, e^{-x^2} \to 0$. Thus, assessing the defnite integral at the bounds, we get

$$\frac{c}{x} e^{-x^2} \Big|_{\sqrt{\gamma \ln(A)}}^{\infty} = P_{\geq}(A)$$

$$= 0 - \frac{c}{\sqrt{\gamma \ln(A)}} e^{-\left(\sqrt{\gamma \ln(A)}\right)^2}$$

$$= -\frac{c}{\sqrt{\gamma \ln(A)}} e^{-\gamma \ln(A)}$$

$$\implies P_{\geq}(A) = -\frac{c A^{-\gamma}}{\sqrt{\gamma \ln(A)}} \tag{1.c.ii}$$

0.2 Problem 2:

For the discrete version of the HOT model, every configuration of a forest has a cost associated with it's particular placement of trees. Since we are modeling the prevention of forest fires, the cost of a particular configuration of trees is proportional to a function of the probability of a fire outbreak and the number of connected components at a particular site on the lattice where a tree exists. The specific relationship can be written more formally as

$$C_{fire} \propto \sum_{i=1}^{N_{sites}} p_i \, a_i \tag{5}$$

where a_i is the area of the i^{th} site's region, and p_i is the average probability of a fire at site i over a given period of time.

Ideally, we want to minimize the cost of a fire outbreak; i.e: we want to maximize the number of trees, while minimizing the effect a fire would have given a large collection of trees that are adjacent (A.K.A a forest). We model this by leaving specific sites on the lattice unoccupied after each time step. These unoccupied sites limit the size of the greatest connected component and mitigate the risk of planting a tree at specific sites. In essence, these sites are firewalls that constrain the cost of a fire, C_{fire}. We write the constraint for building $(d-1)$-dimensional firewalls in d dimensions as

$$C_{firewalls} \propto \sum_{i=1}^{N_{sites}} a_i^{\frac{(d-1)}{d})} = \sum_{i=1}^{N_{sites}} a_i^{-\frac{1}{d}} \tag{6}$$

We understand that because the constraint is what helps minimize C_{fire}, that $C_{fire} \propto C_{firewalls}$. We can employ Lagrange Multipliers as a strategy to minimize C_{fire} subject to the constraint of $c_{firewalls}$, in order to demonstrate this and show how a_i and p_i relate to one another.

Let, $f(p_i, a_i) = C_{fire}$ be constrained under $g(d, a_i) = C_{firewalls}$, then

$$\mathscr{L}(p_i, a_i, \lambda) = f(p_i, a_i) - \lambda\, g(d, a_i)$$

$$\implies \nabla_{a_i} \mathscr{L}(p_i, a_i, \lambda) = \nabla_{a_i} f(p_i, a_i) - \lambda\, \nabla_{a_i} g(d, a_i)$$

$$\implies \nabla_{a_i} \mathscr{L}(p_i, a_i, \lambda) = \nabla_{a_i} C_{fire} - \lambda\, \nabla_{a_i} C_{firewalls}$$

$$\implies 0 = \nabla_{a_i} C_{fire} - \lambda\, \nabla_{a_i} C_{firewalls}$$

$$\implies \nabla_{a_i} C_{fire} = \lambda\, \nabla_{a_i} C_{firewalls}$$

From this, we may say

$$\nabla_{a_i} C_{fire} \propto \nabla_{a_i} C_{firewalls} \tag{7}$$

Differentiating with respect to a_i, we see that

$$\nabla_{a_i} C_{fire} \propto \frac{\partial C_{fire}}{\partial a_i} \sum_{i=1}^{N_{sites}} p_i a_i$$

$$= \frac{\partial}{\partial a_i} \left(p_1 a_1 + p_2 a_2 + p_3 a_3 + \ldots p_i a_i + \ldots \right)$$

$$= \left(\cancel{p_1 a_1} + \cancel{p_2 a_2} + \cancel{p_3 a_3} + \ldots p_i + \ldots \right)$$

$$= p_i \therefore \tag{8}$$

and, similarly

$$\nabla_{a_i} C_{firewalls} \propto \frac{\partial C_{firewalls}}{\partial a_i} \sum_{i=1}^{N_{sites}} a_i^{-\frac{1}{d}}$$

$$= \frac{\partial}{\partial a_i} \left(a_1^{-\frac{1}{d}} + a_2^{-\frac{1}{d}} a_3^{-\frac{1}{d}} + \ldots + a_i^{-\frac{1}{d}} + \ldots \right)$$

$$= \left(\cancel{a_1^{-\frac{1}{d}}} + \cancel{a_2^{-\frac{1}{d}}} \cancel{a_3^{-\frac{1}{d}}} + \ldots + a_i^{-\frac{1}{d} - 1} + \ldots \right)$$

$$= a_i^{-\frac{1}{d} - 1}$$

$$= a_i^{-(1 + \frac{1}{d})} \therefore \tag{9}$$

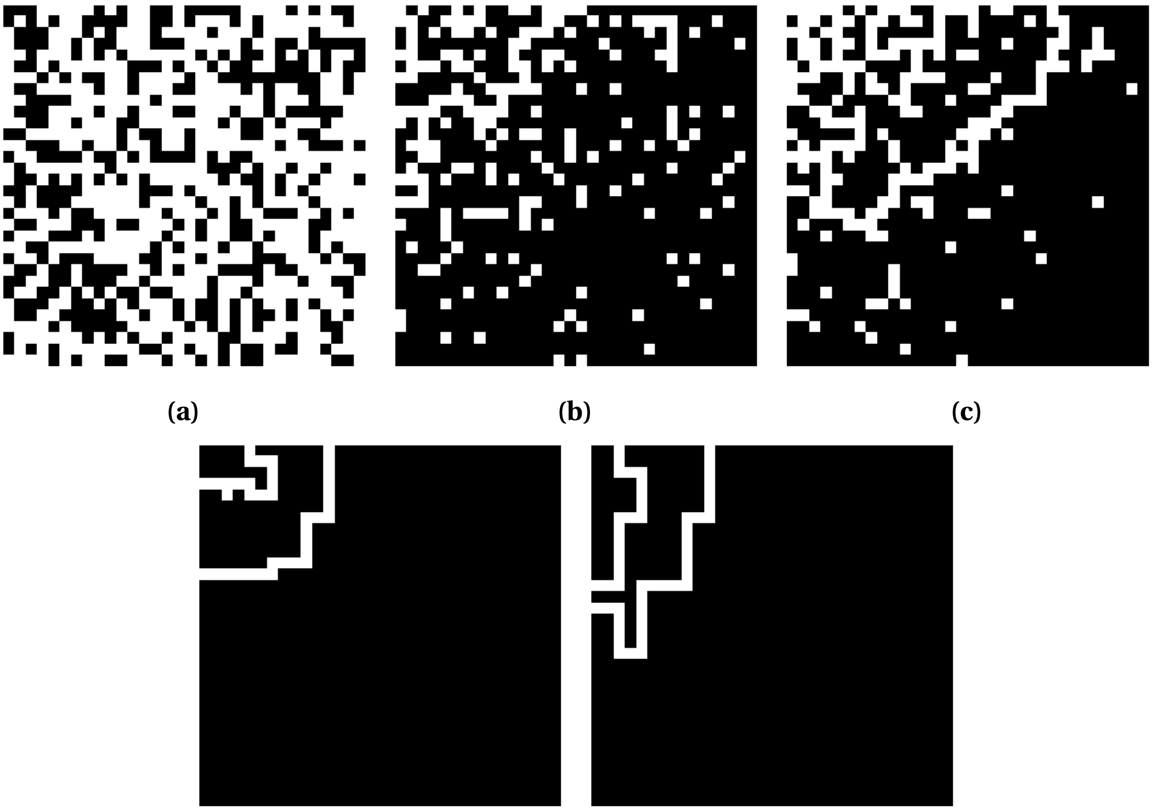
From (**7**), (**8**), and (**9**) we get

$$p_i \propto a_i^{-(1 + \frac{1}{d})} \therefore$$

0.3 Problem 3:

Below is output from a version of the discrete HOT model in 2-d. I took the following approach; create $D = 1, 2, 3, \ldots, L,$ or L^2 2-d arrays of size $L = 32, 64,$ or 128 (I could not manage size 256) sites at each $t = L^2$ time steps. At each of the time steps, I picked from a randomly shuffled list of indices for each of the D 2-d arrays at which to place a new tree (represented by a 1). I would then keep only the lattice configuration with the highest yield given a probability matrix that described the likely hood of a fire outbreak at any site, and a matrix describing the cost of the current configuration of trees. At the passing of the last time step, I had a record of only the configurations that provided the highest yield. I addition, I kept various records of the density, cost, and yield at each time step so as to understand the evolution of the design structure of the 2-d model given D configurations per time step.

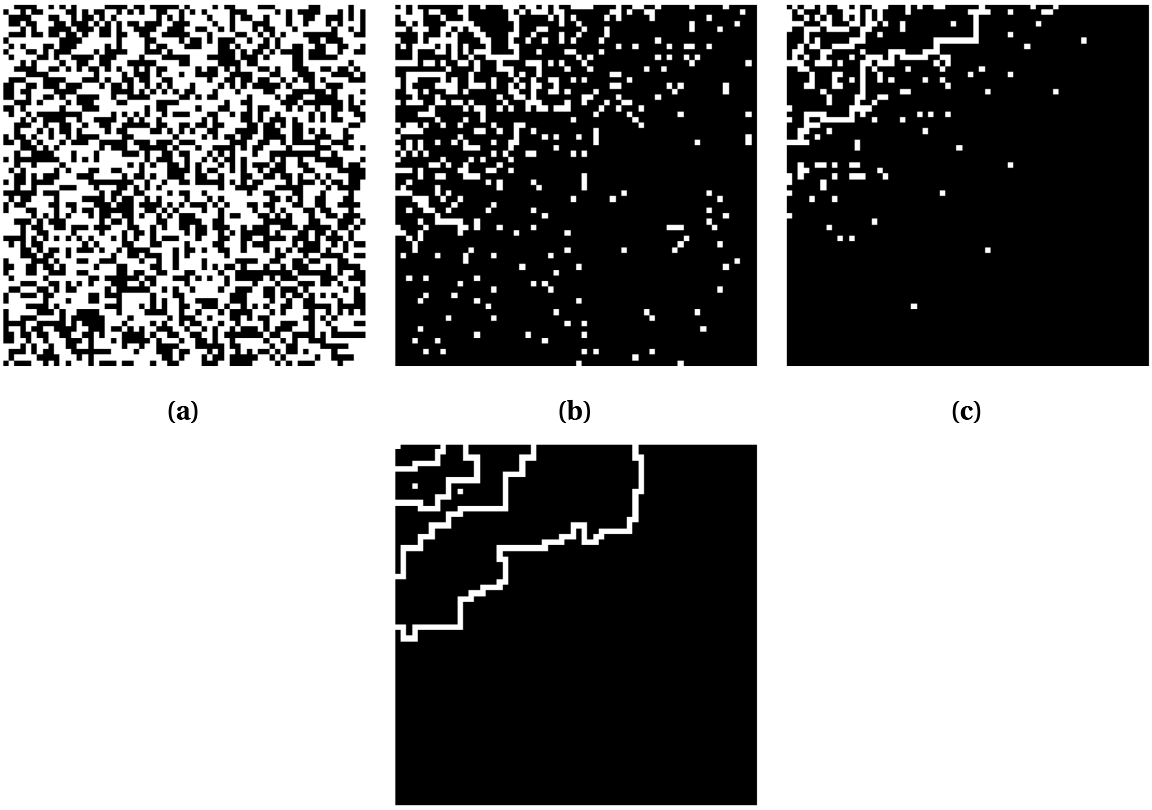
Figures (**4**) through (**6**) show examples of the evolution of the optimal lattice configurations given varying sizes L and D.



(a) (b) (c)

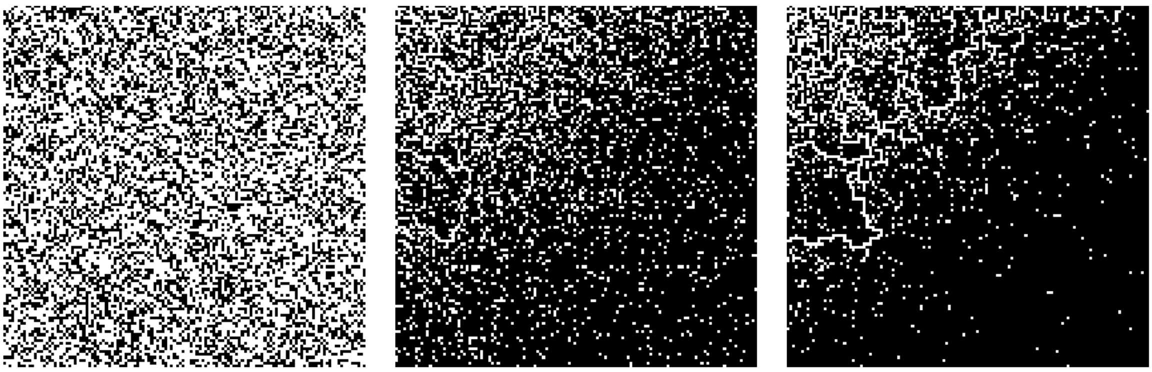
(d) (e)

Figure 4: Peak yields for varying forest densities: From (a) to (e): $D = 1, 2, 3, 32, 1024$, for lattices of size $L = 32$.



(a) (b) (c) (d)

Figure 5: Peak yields for varying forest densities: From (a) to (d): $D = 1, 2, 3, 64$, for lattices of size $L = 64$.

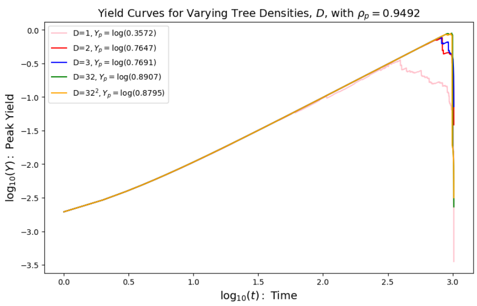


(a) (b) (c)

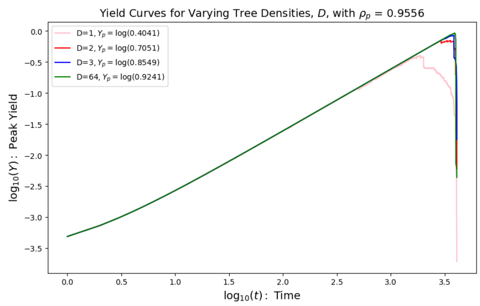
Figure 6: Peak yields for varying forest densities: From (a) to (c): $D = 1, 2, 3$, for lattices of size $L = 128$.

Although the computational cost of creating varying D lattices every time step proved expensive, the figures above clearly highlight how the design of the structure evolves over time under the constraint of the firewalls, which must be constructed to prevent sub-optimal giant components from decreasing the yield potential of a configuration.

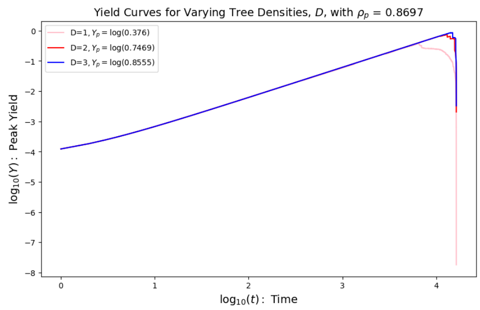
A noteable point to make is that since we plant a tree at a random site at every time step, all lattices will be completely filled at time step L^2. Hence, there will be a configuration representing the most optimal placement of trees. If we plot the yield of a lattice of size L as a function of time, we will see the yield increase for a set number of time steps, and then immediately drop to 0. This phenomenon can be seen in the yield curves in **Figure**(**7**)



(a)



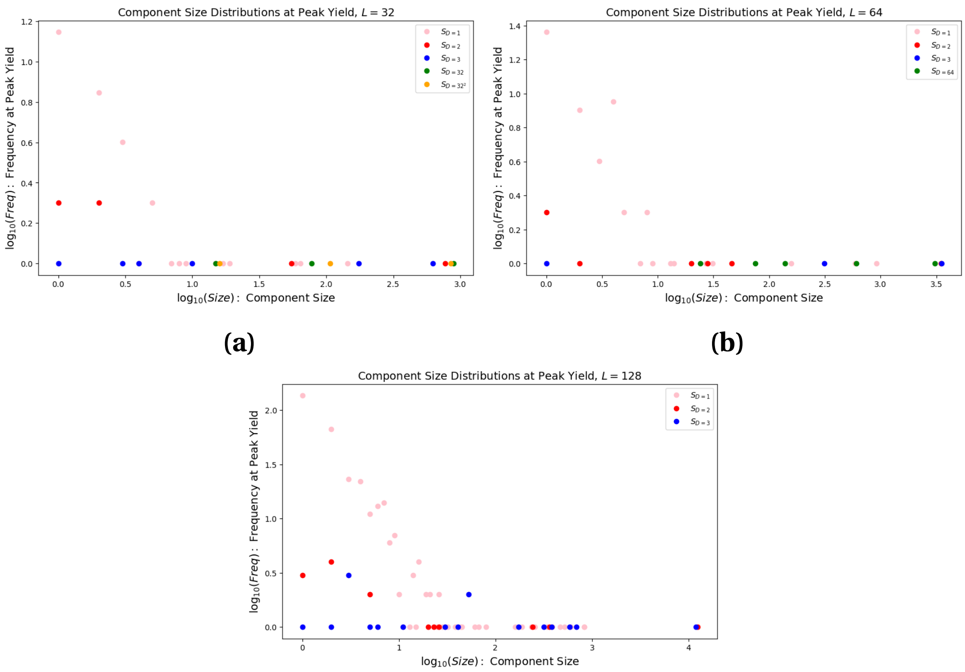
(b)



(c)

Figure 7: Peak yields for varying forest densities over time. (a) to (c): yield curves for lattices of size $L = 32, 64$, and 128 respectively.

In **Figure**(**8**) we show the component size distribution at peak yield for lattices of size L and with varying densities D. Although the plots become more sparse as L increases - a result of computational power - we can clearly see that as D increases, the component size distributions become more sparse and the presence of a power law size distribution of component sizes diminishes.



(a) **(b)** **(c)**

Figure 8: Component size distributions for varying forest densities over time. (a) to (c): component size distributions for lattices of size $L = 32, 64$, and 128 respectively.

As I was only able to compute an output for $D = L^2$ for $L = 32$, **Figure**(**9**) shows only the component size distributions for $L = 32$, $D = 1024$. As described in the paper, and visible in the distribution plot, as the density of the forest increases, the constraint of the firewalls limits the number of components of each size that we see. Essentially, as the design of the configuration becomes more optimal, the presence of a power law size distribution diminishes.

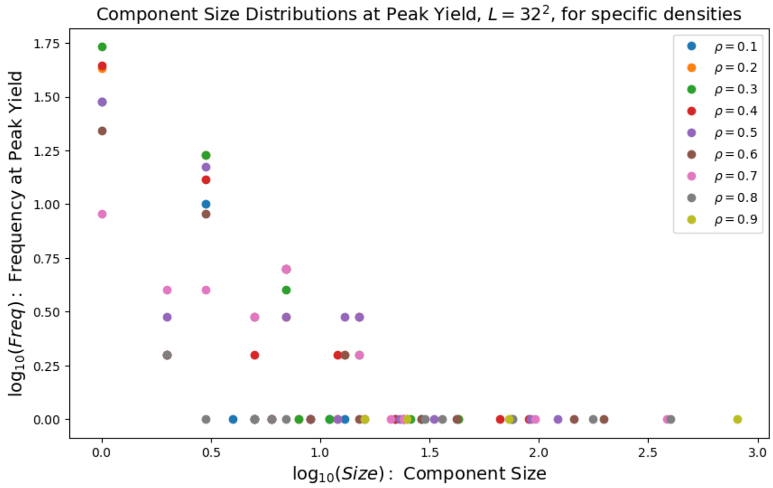


Figure 9: Component size distribution for forests at peak yield, for varying densities. Note that the plot title is slightly off and should read: "Component Size Distributions at Peak Yield, $L = 32$, $D = 32^2$, for Specific Densities."