A Mathematical Model For the Spread of a Virus

N. R. Sheeley, Jr

Alexandria VA November 13, 2020

ABSTRACT

This paper describes a mathematical model for the spread of a virus through an isolated population of a given size. The model uses three, color-coded components, called molecules (red for infected and still contagious, green for infected, but no longer contagious, and blue for uninfected). In retrospect, the model turns out to be a digital analog of the well-known SIR-model of Kermac & McKendrick (1927). In our RGB-model, the accumulated infections go through three phases, beginning at a very low level, then changing to a transition ramp of rapid growth, and ending in a plateau of final values. Consequently, the differential change or growth rate begins at 0, rises to a peak corresponding to the maximum slope of the transition ramp, and then falls back to 0 in a time comparable to the time to reach the peak. The properties of these time variations, including the slope, duration, and height of the transition ramp, and the width and height of the infection rate, depend on a single parameter. This parameter is the time that a red molecule is contagious divided by the average time between collisions or encounters of the molecules. Various milestones, including the starting time of the transition ramp, the time that the accumulating number of infections obtains its maximum slope, and the location of the peak of the infection rate all depend on the size of the population in addition to the contagious lifetime ratio. Explicit formulas for these quantities are derived and summarized.

Contents

1 Introduction **4**

2 The Model **4**

3 Permanently Contagious Red Molecules **5**

3.1 Deriving the Infection-Spreading Equation 5

3.2 Plotting the Spread 6

3.3 Analytical Solution of the Difference Equation 8

4 The Spreading When Red Molecules Have a Finite Lifetime **9**

4.1 The Model 9

4.2 The Equations 10

4.3 The Plots 12

5 Quantitative Results: Profile Dependence on N_0 and c **15**

5.1 The Final Value of the Growth Curve 15

5.2 The Height and Width of the Growth-Rate Peak 17

5.3 Location of the Peak of the Growth Rate 21

5.3.1 The slope-intercept method 22

5.3.2 The point-slope method 25

6 Red, Green, and Blue Molecules **27**

6.1 Contributions of the red and green molecules 27

6.2 Herd 'Immunity' 30

7 Changing Social Distancing During an Epidemic **37**

7.1 Removing the Social Distancing Entirely 37

7.2 Sudden Increase of **c** 38

7.3 Sudden Decrease of **c** 40

8 Summary **41**

9 Discussion **46**

A Solution For $\nu_{\mathrm{j}} << 1$ **50**

B Estimating the Starting Point of the Region of Rapid Increase **53**

C Approximating the Linearized-Solution Parameters, σ_{c} and ρ_{c} **54**

D Comparison with the Solutions of the SIR Equations **57**

1. Introduction

The spread of the corona virus and its infection as covid-19 has become a central issue in all of our lives. Curves showing the accumulated infections as well as the infection rate occur in the news every day. Most of us are 'social distancing' in order to 'flatten the curve' and we are receiving a variety of advice about how long it will take before it is safe to return to our normal activities. Consequently, I thought it would be both interesting and useful to calculate these curves using a simple model for the infection. It would be interesting just to see if I could actually do the calculation, and it would be useful if the solution made sense and gave me an appreciation of how the spreading takes place. The idea is not to perform a fully realistic calculation, but to learn what the relevant parameters are and to gain some insight about how they affect the spread of the disease. This paper is a summary of what I have learned.

2. The Model

Let us regard the affected population as a gas of colliding 'molecules' of a given number, N_0, not nearly as big as Avogadro's number, $N_0 = 6 \times 10^{23}$ atoms/mole, but $\sim 10^3 - 10^8$, comparable to populations of various regions of the country and the world. So we imagine that a population is like a gas of N_0 molecules, colliding with each other in a region of constant volume. Presumably, this gas will have characteristics of a real gas like temperature, pressure, and mean free path that can be determined. However, for the moment, we will bypass this interesting question and move forward to calculate the spread of the infection.

We assume the problem starts with $N_0 - 1$ molecules that are uninfected and one molecule that is both infected and contagious, in the sense that it will infect another molecule when it collides with that molecule. For simplicity, we suppose that there is an average time, τ_s, between collisions, which we will use as the step time in our calculation. Referring to the left panel of Figure 1, we begin at step 0 with the index $j = 0$ when a contagious (red) molecule is inserted into the gas of $N_0 - 1$ uninfected (blue) molecules. By the next step in the right panel of Figure 1, the red molecule has collided with a blue one, making it red, so that two red molecules can go on to infect two other blue molecules in the next collision.

An important aspect of this problem is the finite time that a red molecule remains contagious. For the corona virus, we have heard that the lifetime is on the order of 2 weeks, the time that some potentially infected people have been told to quarantine. However, we will delay our treatment of this effect until later in this paper when we will have gained some insight and experience by solving the simpler problem for a permanently contagious virus.

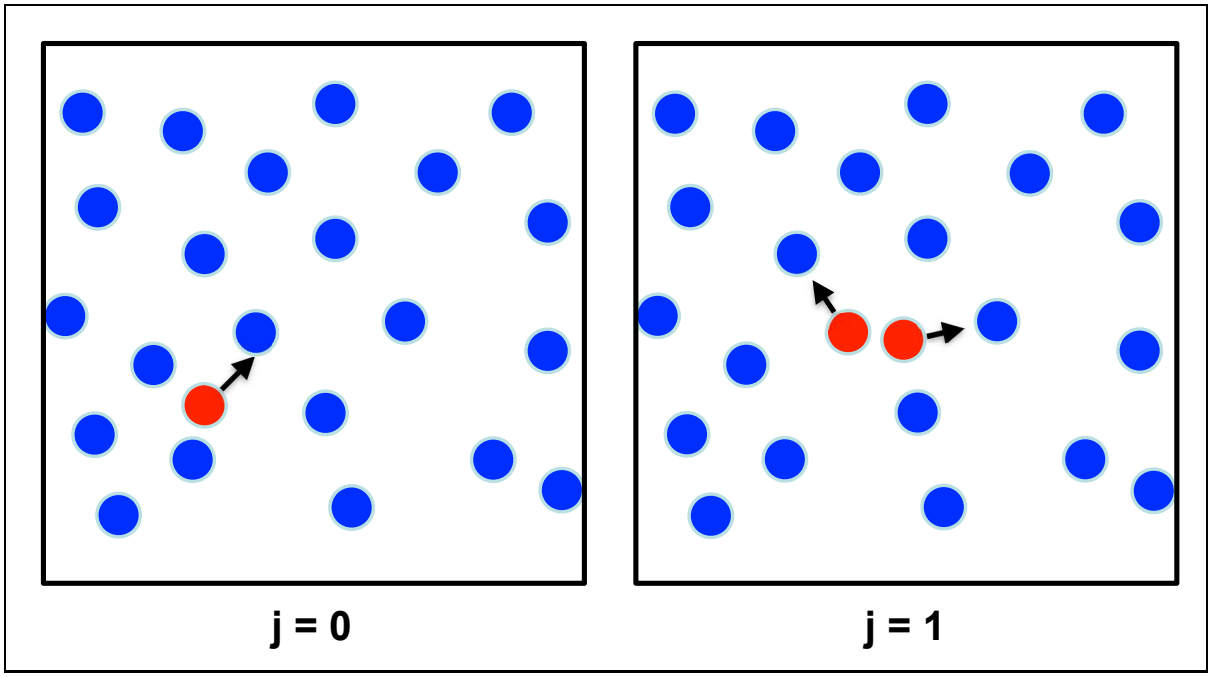

Fig. 1.— (left) A gas of uninfected (blue) molecules with a contagious (red) molecule inserted at the initial time $j = 0$. (right) The same gas after one collision time when the red molecule has encountered a blue molecule and infected it (*i.e.* changed it to red).

3. Permanently Contagious Red Molecules

3.1. Deriving the Infection-Spreading Equation

Let us assume that the problem has advanced to the $j-1$ step at which point the number of infected molecules is represented by the symbol, $n_{inf}^{(j-1)}$, where the exponent, $j-1$, is a tag and not a power to which n_{inf} is to be raised. The question is then what is the number of infected molecules in the next step, as represented by the symbol $n_{inf}^{(j)}$? For the first few steps with $j \sim 1$, the probability, p_{j-1}, of colliding with an uninfected molecule will be close to 1 so that $n_{inf}^{(j)}$ would be $2n_{inf}^{(j-1)}$ (*i.e.* just twice the present number of infected molecules before the collision). However, for larger values of j, we must consider the case of $p_{j-1} \neq 1$.

The probability that a red molecule will collide with another infected molecule is just $(n_{inf}^{(j-1)} - 1)/(N_0 - 1)$, so the probability that it will collide with an uninfected blue molecule is

$$p_{j-1} \;=\; 1 - \frac{n_{inf}^{(j-1)} - 1}{N_0 - 1}. \tag{3.1}$$

Consequently,

$$n_{inf}^{(j)} \;=\; n_{inf}^{(j-1)} \;+\; n_{inf}^{(j-1)} p_{j-1}, \tag{3.2}$$

and therefore

$$n_{inf}^{(j)} \;=\; n_{inf}^{(j-1)} \;+\; n_{inf}^{(j-1)} \left\{ 1 - \frac{n_{inf}^{(j-1)} - 1}{N_0 - 1} \right\}. \tag{3.3}$$

Now, for simplicity, we normalize the number of molecules to the total number, N_0, in the gas, and define $\nu_j \;=\; n_{inf}^{(j)}/N_0$ as the normalized quantity. Thus, $\nu_0 \;= 1/N_0$, and

$$\nu_j \;=\; \nu_{j-1} \;+\; (\frac{1}{1-\nu_0})\nu_{j-1}(1-\nu_{j-1}) \;=\; \nu_{j-1}\left\{ 1 + (\frac{1}{1-\nu_0})(1-\nu_{j-1}) \right\}. \tag{3.4}$$

3.2. Plotting the Spread

Next, in Figure 2, we plot the growth curve, ν_j, and the growth rate, $\Delta\nu_j = \nu_j - \nu_{j-1}$, versus the number of steps (j) or collision times ($j\tau_s$) that have elapsed since the start of the infection. The difference equation Eq.(3.4) gives the sequence of points that are plotted as red dots. (The continuous blue curves are plotted from the analytical solution to Eq. (3.4) derived in the next subsection.) As one can see, for this population of $N_0 = 1 \times 10^5$, the number of infections reaches approximately one half of its final value after about 16 collisions,

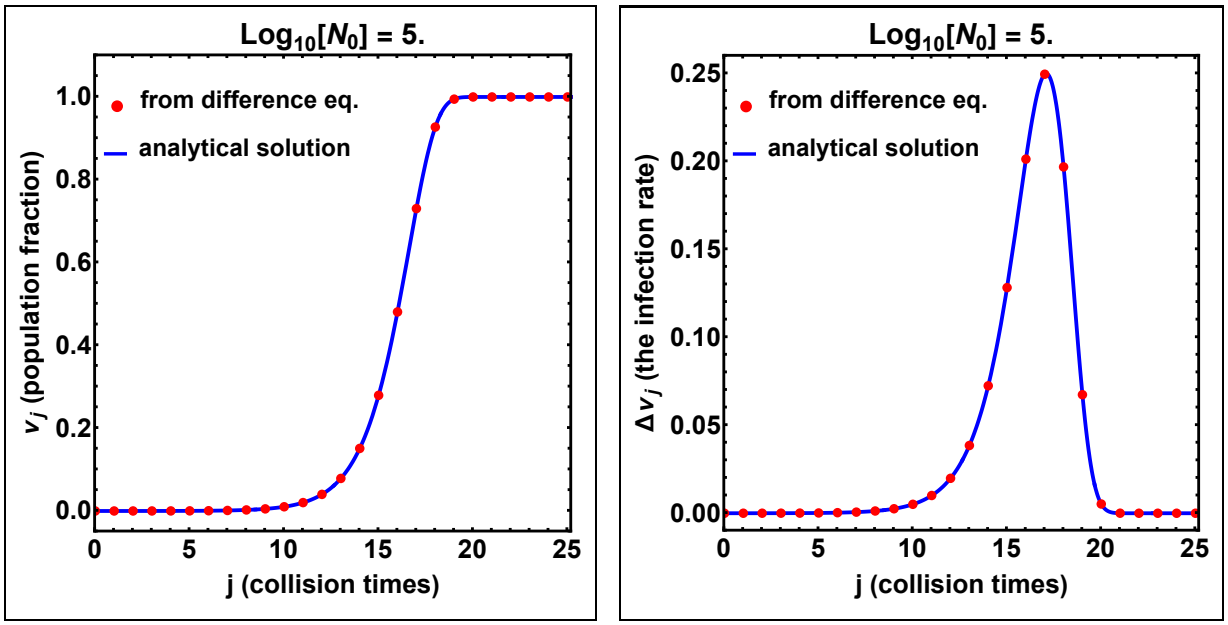

Fig. 2.— (left): ν_j from Eq.(3.4) (red points) and Eq.(3.13a) (blue curve); (right): $\Delta\nu_j$ from Eq.(3.11) (red points) and Eq.(3.13b) (blue curve).

which is one collision before the growth rate in the right panel reaches its peak value of 0.25 at $j_{max} \approx 17$. At this point, 1/4 of the total population is being infected per collision time. The full width at half maximum is approximately 4 collision times, which means that the area under the curve is about 1, consistent with the final height of the infection curve in the

left panel. Also, note that each profile is asymmetric. In the left panel, the trip from the point of maximum slope to the upper level is shorter than the trip from the lower level to the point of maximum slope. Likewise, in the right panel, the trip from the peak to the final base level is shorter than the trip from the initial base level to the peak.

Although the shape of the growth-rate profile does not vary with N_0, its location, j_{max}, does vary with N_0, as shown by the measurements that are plotted as red dots in Figure 3. The least-squares best fit to these measurements is shown by the dashed red line, whose

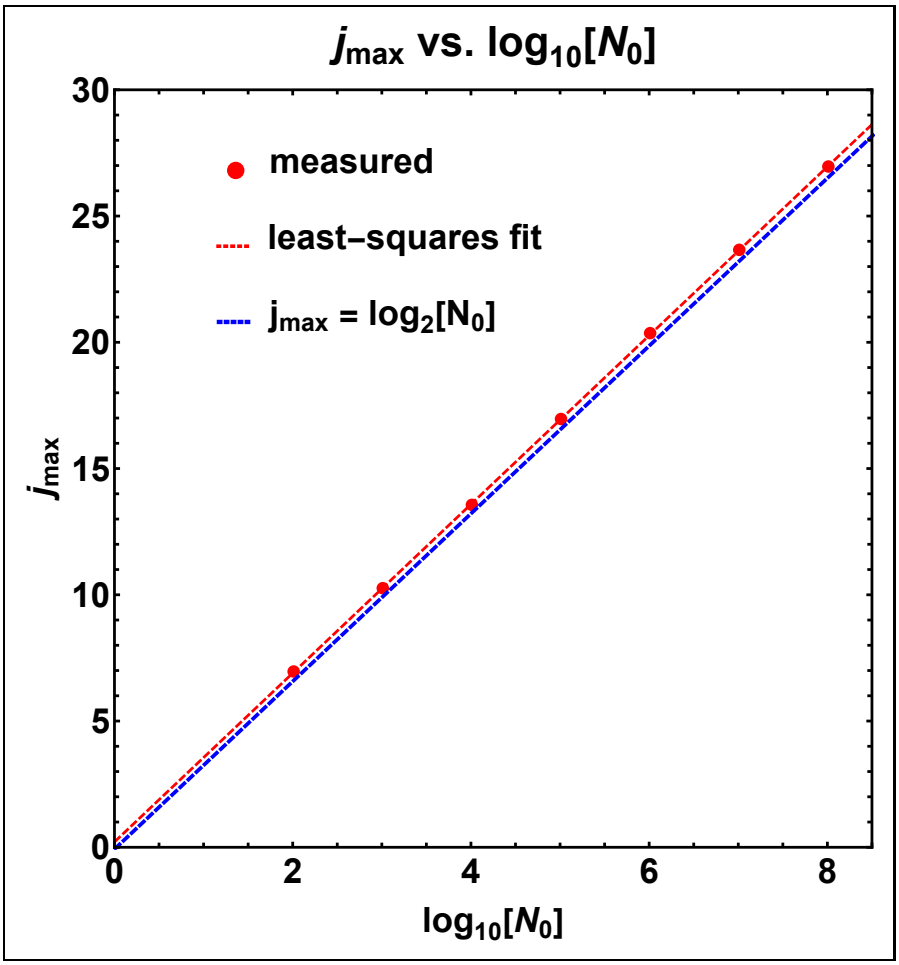

Fig. 3.— Measurements of the peak location, j_{max}, plotted versus $\log_{10} N_0$ (red dots), and their least-squares linear fit (dashed red line), compared with the relation, $j_{max} = \log_2 N_0 \approx 3.322 \log_{10} N_0$ shown by the dashed blue line.

equation is

$$j_{max} \;=\; 0.286 \;+\; 3.343 \log_{10} N_0. \tag{3.5}$$

This red line is nearly coincident with the dashed blue line, which represents the relation

$$j_{max} \;=\; \log_2 N_0 \;=\; \frac{\log_{10} N_0}{\log_{10} 2} \approx 3.322 \;\log_{10} N_0, \tag{3.6}$$

which is the number of two-factor collisions that it takes to reach the total population, N_0.

3.3. Analytical Solution of the Difference Equation

It is instructive to see if we can find an analytical solution to Eq. (3.4). Let's begin with the growth equation in the form

$$\nu_j \;=\; \nu_{j-1}\left\{1+(\frac{1}{1-\nu_0})(1-\nu_{j-1})\right\} \;\approx\; \nu_{j-1}(2-\nu_{j-1}). \tag{3.7}$$

Initially, $\nu_{j-1} << 1$, so that $\nu_j \approx 2\nu_{j-1} = 2^2\nu_{j-2} = \cdots = 2^n\nu_{j-n}$. Setting $j = n$, we obtain the solution $\nu_n = 2^n\nu_0$, corresponding to an exponentially increasing number of infections. So it is the second term that limits the exponential growth and stops the infection.

We can use a similar approach when the second term is included. In this case, we change variables by introducing $\mu_j = 1 - \nu_j$ and obtain

$$1-\mu_j \;=\; (1-\mu_{j-1})[2-(1-\mu_{j-1})] \;=\; (1-\mu_{j-1})(1+\mu_{j-1}) \;=\; 1-\mu_{j-1}^2. \tag{3.8}$$

Consequently, $\mu_j = \mu_{j-1}^2 = \mu_{j-2}^4 = \mu_{j-3}^8 = \cdots = \mu_{j-n}^{2^n}$. Setting $j = n$, we obtain $\mu_n = \mu_0^{2^n}$. Finally, changing back to ν_j, we have

$$\nu_j \;=\; 1-(1-\nu_0)^{2^j} \tag{3.9}$$

as the solution to Eq.(3.7).

Next, noting that $1 - \nu_0 \approx e^{-\nu_0}$, we can change Eq.(3.9) to

$$\nu_j \approx 1-e^{-\nu_0 2^j} \;=\; 1-2^{-(\nu_0/ln2)2^j} \;=\; 1-2^{-2^{(j-k)}}, \tag{3.10}$$

where $(\nu_0/ln2) = 2^{-k}$. Because $\nu_0 = 1/N_0$, it follows that $k = \log_2(N_0 ln2)$, which is 1 unit less than j_{max}, the location of the peak of the growth rate, $\Delta\nu_j = \nu_j - \nu_{j-1}$, as we shall see next.

We begin by using Eq.(3.7) to evaluate $\Delta\nu_j = \nu_j - \nu_{j-1}$. The result is

$$\Delta\nu_j \;=\; (\frac{1}{1-\nu_0})\nu_{j-1}(1-\nu_{j-1}) \approx \nu_{j-1}(1-\nu_{j-1}). \tag{3.11}$$

The quantity, $\nu_{j-1}(1-\nu_{j-1})$, has its maximum value of 1/4 when $\nu_{j-1} = 1/2$. Thus, $\Delta\nu_j$ has its maximum value when $\nu_{j-1} = 2^{-2^{j-1-k}} = 1/2$. This happens when $j - 1 - k = 0$, and therefore

$$j_{max} \;=\; j \;=\; k+1 \;=\; 1 \;+\; \log_2(N_0 ln2), \tag{3.12}$$

which is $j_{max} = \log_2 N_0 + 0.471$ when the numerical value of $1 + \log_2(ln2)$ is used. Also, because $\nu_{j-1} = (1-\nu_0)^{2^{j-1}} \approx 2^{-2^{(j-j_{max})}}$, it follows that $\Delta\nu_j = 2^{-2^{(j-j_{max})}} \left\{1 - 2^{-2^{(j-j_{max})}}\right\}$. Summarizing the results for both ν_j and $\Delta\nu_j$, we have

$$\nu_j \;=\; 1 \;-\; 2^{-2^{\{j-(j_{max}-1)\}}}, \tag{3.13a}$$

$$\Delta\nu_j \;=\; 2^{-2^{(j-j_{max})}} \left\{1 - 2^{-2^{(j-j_{max})}}\right\}. \tag{3.13b}$$

As we saw in Figure 2, these solutions for ν_j and $\Delta\nu_j$ agree very well with the data points that were obtained from the recursive equations, Eq.(3.4) (or (3.7)) and Eq.(3.11). The point at $j = 17$ occurred very slightly to the left of the peak at $j_{max} = 17.1$. Also, we previously noted in the right panel of Figure 2 that the full-width at half maximum (w) of the growth-rate peak is approximately 4 collision times. Now, with Eq.(3.10), we can calculate this width precisely by setting $\nu_{j-1}(1-\nu_{j-1}) = (1/2)\Delta\nu_{max} = 1/8$. This is a quadratic equation whose roots are $\nu_{j-1} = (1/2)(1 \pm 1/\sqrt{2})$. Setting them equal to $2^{-2^{(j-k)}}$, we obtain $2^{j-k} = 1 - \log_2(1 \pm 1/\sqrt{2})$, whose numerical values are 0.2284 and 2.7716. The corresponding values of $j-k$ are -2.130 and 1.471, and their difference (the full-width at half maximum) is $w = 3.601$, about 10% less than our original estimate from Figure 2.

It is interesting to compare the width and height of this growth-rate curve with the total area under it (which we expect to be 1 because $\sum_{j=0}^{j=25} \Delta\nu_j = \nu_{25} - \nu_0 = 1 - 1/N_0 \approx 1$). For our calculated full-width at half maximum, the product of height and width is $0.25\times3.601 = 0.900$, which is slightly less than 1. On the other hand, if we had used the full-width at $1/e$ maximum instead of the full-width at half maximum, we would have obtained $0.25\times4.398 = 1.100$, which is slightly greater than 1. It is an interesting coincidence that the average of these two widths is almost exactly 4.000, which gives an area of 1.000. Now, let us return to the more realistic case in which the red molecules are contagious for a finite time.

4. The Spreading When Red Molecules Have a Finite Lifetime

4.1. The Model

In this section, we shall assume that the infected red molecules do not remain contagious indefinitely, but lose their ability to infect other molecules after a finite time, c. We suppose that $c > 1$, so that the red molecules remain contagious for at least one collision time. Otherwise, the molecules would lose their ability to infect before they encounter uninfected blue molecules in the next collision, and the epidemic would be over. So, we suppose that after a red molecule suffers c collisions, it loses is ability to infect others. At this point, we change its color from red to green.

In the real world, a molecule could become green either because it recovered from the spreading disease or because it did not recover and died. In the latter case, the green molecule ought to be removed from the gas. For the current corona virus, this applies to about 3-5% of the green molecules (or less depending on how many of the infections were not reported). Consequently, in this paper we will neglect this effect and keep all of the green molecules in the gas. As we shall see later, these green molecules provide a kind of 'shielding effect' by reducing the probability that the red molecules will come in contact with the remaining blue molecules during the next collision.

Figure 4 shows the progress for $c = 2$. In the left panel, one collision has taken place, and the original red molecule from $j = 0$ has infected another molecule at $j = 1$. These two red molecules continue on to infect two more molecules by $j = 2$ (right panel), but by this

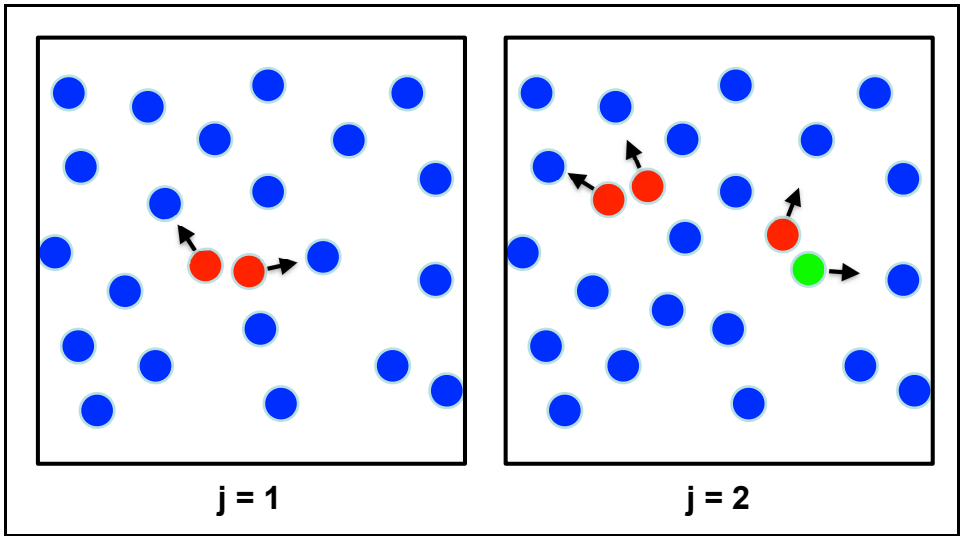

Fig. 4.— Same as Figure 1, but with a finite contagious time of $c = 2$ collision times. After 2 collision times, the initially infected molecule (now green) is no longer contagious.

time, the original red molecule has lost its contagiousness and has become green. Like the three red molecules, the green molecule will go on to collide with other blue molecules, but it will not infect them or change their color.

4.2. The Equations

Proceeding as we did in the previous section, the probability that a red molecule will collide with a blue molecule is still given by Eq. (3.1). However, because some of the infected molecules are no longer red (*i.e.* contagious), Eq. (3.2) must be replaced by

$$n_{inf}^{(j)} = n_{inf}^{(j-1)} + n_c^{(j-1)} p_{j-1}, \tag{4.1}$$

where $n_c^{(j-1)}$ is the number of contagious molecules at the time $j-1$. Now, we need to find an equation for the number of contagious molecules.

To get a feel for this process, we refer to Figure 5, which is a sketch of the evolving number of red and green molecules as a function of the collision time, j, assuming that there

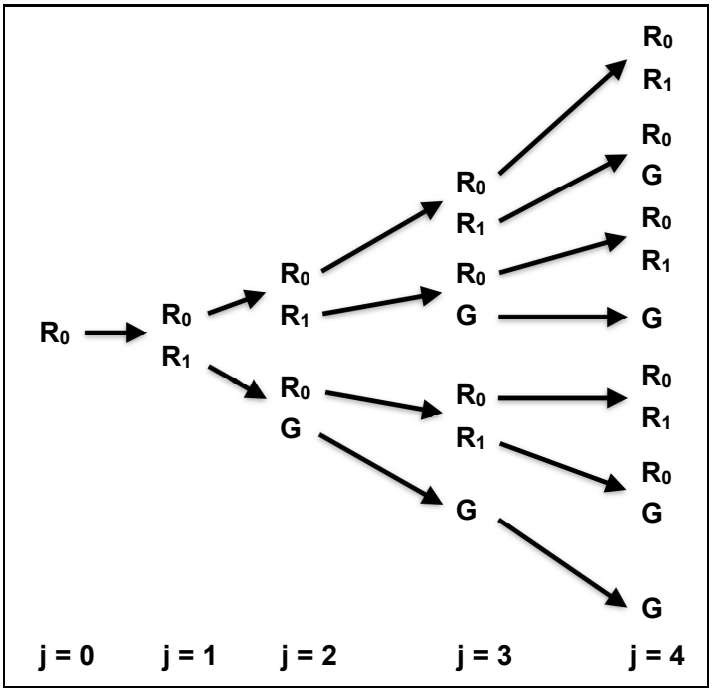

Fig. 5.— The spread of infection during $j = 0-4$, assuming that infected molecules remain contagious for 2 collision times ($c = 2$), and are then colored green (G). At a time, j, the number of contagious molecules is equal to the number of infected molecules at that time minus the number of infected molecules 2 steps earlier. The subscript on R refers to the time since that molecule became infected.

are no collisions between red molecules during this time. At the left end of the array, we begin at $j = 0$ with R_0, which is a symbol for a single red molecule of 'age' 0. After the first collision at $j = 1$, there are two red molecules - the original red molecule, R_1, whose age is now 1, and the newly infected molecule, R_0. One step later, there are 4 infected molecules - R_0 and R_1, derived from R_0 at time $j = 1$, plus R_0 and G, derived from R_1 at time $j = 1$. Instead of calling the the originally infected molecule by the name R_2, we realize that this molecule is not contagious anymore, and we represent it by the letter G. As time passes to $j = 3$ and $j = 4$, the pattern expands, with each R_0 always producing an R_0 and an R_1, with each R_1 producing an R_0 and a G, and with each G always remaining a G.

From the evolving pattern in Figure 5 with $c = 2$, we can see that the number of contagious molecules at time $j-1$ is equal to the total number of infected molecules (contagious plus noncontagious) at that time minus the total number of infected molecules 2 steps earlier.

More generally, we can appreciate that all of the contagious molecules (R's) at any time, $j-1$, will lose their contagiousness (and become G's) after an additional time, c, so that

$$n_c^{(j-1)} \;=\; n_{inf}^{(j-1)} \;-\; n_{inf}^{(j-1-c)}, \tag{4.2}$$

even when there are collisions between red molecules. Combining Eqs. (3.1), (4.1), and (4.2), and normalizing to the total population, N_0, we have

$$\nu_j \;=\; \nu_{j-1} \;+\; \left(\frac{1}{1-\nu_0}\right)(\nu_{j-1}-\nu_{j-1-c})(1-\nu_{j-1}), \tag{4.3}$$

which reduces to Eq. (3.7) when $\nu_{j-1-c}=0$. Likewise, the difference form of Eq. (4.3) is

$$\Delta\nu_j \;=\; \left(\frac{1}{1-\nu_0}\right)(\nu_{j-1}-\nu_{j-1-c})(1-\nu_{j-1}), \tag{4.4}$$

which reduces to Eq. (3.11) when $\nu_{j-1-c}=0$.

4.3. The Plots

Our next step is to plot ν_j and $\Delta\nu_j$ versus j for various values of the parameter, c. To do this, we set $\nu_0 = 1/N_0$ and use Eqs. (4.3) and (4.4) for $j \;\geq\; 1$. Also, we add the condition that $\nu_j = 0$ for $j < 0$. This means that ν_{j-1-c} will not contribute until $j-1=c$.

Finally, because the influence of the finite contagious time becomes greatest as that time approaches the collision time (*i.e.* as $c \to 1$), we need to find a way to make these equations apply for non-integral values of c between 1 and 2. For this purpose, we used a linear interpolation between values of c in the range $[1,2]$, or more generally in the range $[c_0, c_0+1]$, where c_0 is an integer. Consequently, we replaced ν_{j-1-c} by the expression $\lambda\nu_{j-1-c_0-1} + (1-\lambda)\nu_{j-1-c_0}$, where λ lies in the range $[0,1]$ and $c = c_0 + \lambda$.

Figures 6 and 7 show ν_j and $\Delta\nu_j$ plotted versus j for parametric values of c from 1.25 to 10.0. I chose 10 as the largest contagious time because the curves for $c=10$ and a much larger value of c like $c=100$ were indistinguishable. Thus, once the contagious time reaches 10 collision times, the result is the same as if the red molecules remain contagious for the duration of the problem. (The curves were almost indistinguishable for $c=6$, except that the peak height for the differential curve was 0.24 instead of 0.25.) Consequently, the curves with $c = 10$ are an accurate representation of the result when the red molecules remain contagious for the duration of the problem (*i.e.* the value of ν_j eventually reaches 1, and the height of the peak of $\Delta\nu_j$ is 0.25 at the location where $j_{max} \approx 17$, which is approximately $\log_2 N_0 = 16.6$.)

In retrospect, it is obvious why the growth curve with $c=10$ is indistinguishable from the corresponding curve for permanently contagious molecules (with $c=\infty$). For $c=10$,

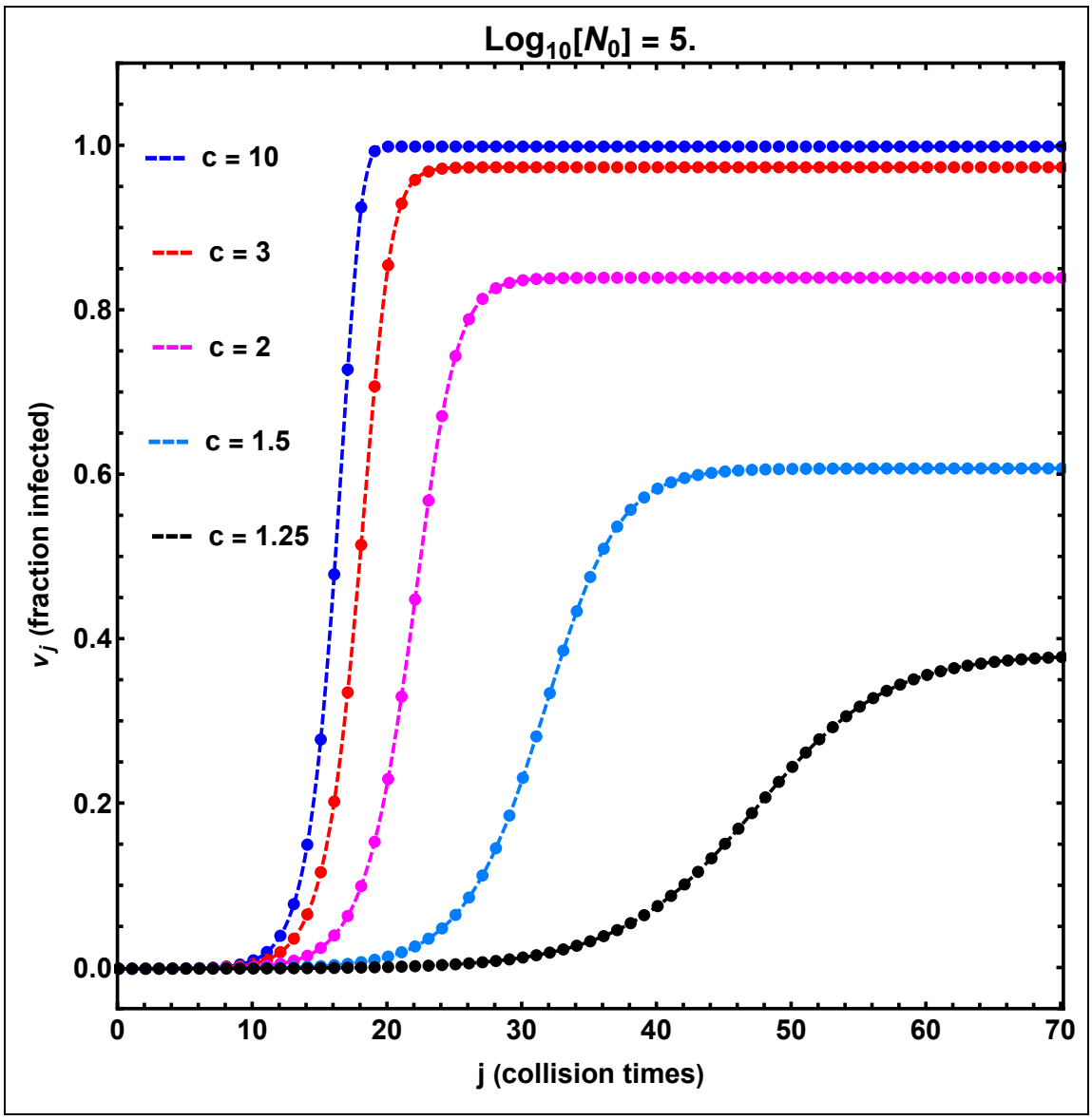

Fig. 6.— The fraction of infected molecules, ν_j, as a function of time for values of the contagious time, c, from 10 to 1.25, showing the delayed onset and the reduced final fraction as the contagious time becomes closer to the collision time.

the lag is so long that the plot of ν_{j-1} finished its rise from 0 to 1 before the lagged quantity, ν_{j-1-c}, became appreciable ($\nu_9 = 0.005$). Equivalently, there are 10 points between $\nu_j \approx 0$ and $\nu_j \approx 1$. For a smaller value of c, the lag is shorter, and ν_{j-1-c} becomes appreciable before ν_j reaches the end of the rising portion of its curve. In addition, there are more points between the beginning and end of the rising portion of the curve. This produces less spacing between the points of ν_j in Figure 6 and lower peaks in the plots of $\Delta\nu_j$ in Figure 7.

For $c = 10$, the fraction of infected molecules reaches 1 at $j \approx 20$ and remains there

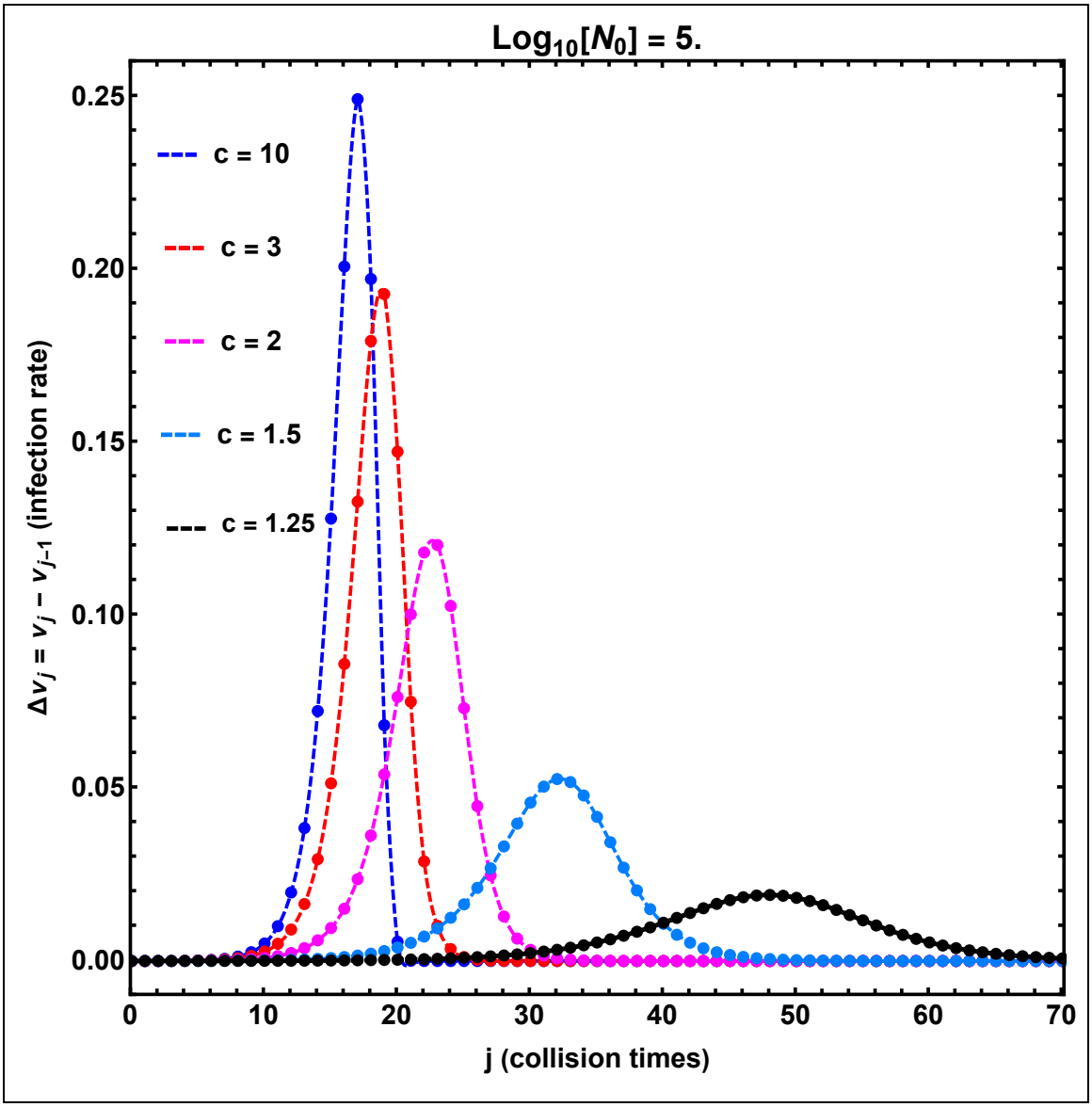

Fig. 7.— The differential fraction of infected molecules, $\Delta\nu_j$, as a function of time, showing that the profiles become lower, wider, and shifted to later times as the contagious time, c, becomes closer to the collision time.

forever. This means that everyone in the population eventually becomes infected. However, as c decreases, the final fraction of infected molecules decreases, reaching about 0.38 when $c = 1.25$ collision times. If the contagious time were about 14 days, $c = 1.25$ would correspond to a collision time of 11.2 days. This means that to reduce the number of infected molecules to 38% of the total population, one would have to increase the average time between collisions to about 11 days. Such 'social distancing' prevents 62% of the population from becoming infected and shows the importance of making the time between collisions as close as possible to the 14-day contagious time. But even for 11 days, the end of the epidemic would be

delayed appreciably from $j \sim 20$ to $j \sim 70$, which is roughly from 1 year to 3 years if the collision time is 2 weeks.

Figure 7 shows that the infection rate weakens, broadens, and shifts to later times as the contagious time, c, approaches the collision time. I think this is what the media reports for the corona virus are referring to when they use the term 'flattening the curve'. Corresponding plots for other values of N_0 show essentially the same profiles, shifted in time. In the next section, we shall analyze the dependences of these profiles on N_0 and c.

5. Quantitative Results: Profile Dependence on N_0 and c

In the previous section, we saw that the growth rate gradually rises to a peak $\sim 0.25 N_0 {\tau_s}^{-1}$ when the contagious time, c, is large compared to 1, and that the peak height becomes smaller as c approaches 1. In this section, we will determine the location, height, and width of this peak as a function of N_0 and c. Also, we will determine the dependence of the final value of ν_j on these parameters.

5.1. The Final Value of the Growth Curve

We can obtain a clue to the relation between c and ν_f by examining the behavior of ν_{j-1} in the region where the growth curve has its maximum slope. In this region, the slope is fairly constant so that it will have approximately the same value at several neighboring points. In particular, we would expect that

$$\frac{\nu_{j-1} - \nu_{j-1-c}}{c} \approx \nu_{j-1} - \nu_{j-2} \approx \nu_j - \nu_{j-1} \;=\; \Delta\nu_j, \tag{5.1}$$

if c is close to 1. In this case, Eq. (4.4) reduces to

$$\Delta\nu_j \approx c\Delta\nu_j(1 - \nu_{j-1}), \tag{5.2}$$

which would be an identity if $c(1 - \nu_{j-1}) = 1$. Equivalently,

$$\nu_{j-1} \approx 1 - \frac{1}{c}. \tag{5.3}$$

This point of maximum slope lies somewhere between the starting and ending values ν_0 and ν_f, perhaps becoming centered between them if the growth-rate curves become symmetric when c approaches 1 (as Figure 7 suggests). In this case, we would expect that

$$\nu_f \approx \alpha(1 - \frac{1}{c}), \tag{5.4}$$

where α is a number on the order of 2.

Although Eq.(5.4) seems to reproduce the calculated values of ν_f fairly well as c approaches 1, it does not satisfy the requirement that $\nu_f \to 1$ as $c \to \infty$. To satisfy this requirement and still retain the behavior for $c \approx 1$, we can use an exponential of the form $1 - e^{-\alpha(c-1)}$. This expression approaches 1 as $c - 1 \to \infty$, and it approaches $\alpha(c-1)$ as $c - 1 \to 0$. In fact, if we retain only the first-order terms in $c - 1$, we obtain the equality:

$$\alpha(1 - \frac{1}{c}) \;=\; \alpha\left\{\frac{c-1}{1+(c-1)}\right\} \;= \alpha(c-1) \;=\; 1 - e^{-\alpha(c-1)}. \tag{5.5}$$

Figure 8 shows ν_f, computed from Eq.(4.3) and plotted for values of c in the range $1.05 - 4$. The value of $N_0 = 10^5$ used in these calculations determines only the value of j

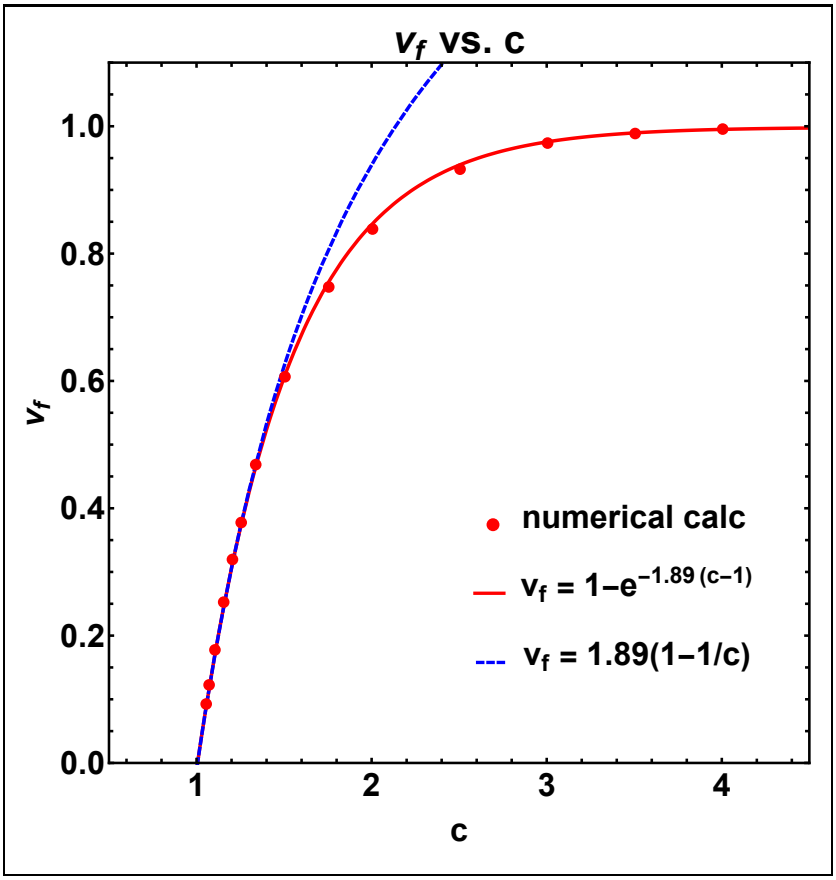

Fig. 8.— The final value, ν_f, of the growth profile plotted versus the contagious lifetime, c, with the solid red line indicating the root-mean-square best fit given by Eq. (5.6) and the dashed blue line indicating the approximation given by Eq. (5.4) with $\alpha = 1.890$.

where the final value of ν_f occurs, but it does not determine ν_f itself. In other words, we would obtain the same plot of ν_f versus c for other values of N_0. The solid red curve is the least-squares-best-fit to these points using the relation

$$\nu_f \;=\; 1 \;-\; e^{-1.890(c-1)}, \tag{5.6}$$

and the dashed blue curve is the approximation given by Eq. (5.4) with $\alpha = 1.890$.

As we can see, Eq. (5.4), represented by the dashed blue line, fits the data very well for c up to 1.5, but the exponential relation given by Eq. (5.6) is required to extend the fit to larger values of c. Consequently, Eq. (5.6) provides the desired relation between the amount of social distancing and the fraction of molecules that eventually become infected (and therefore the fraction that ultimately escapes the epidemic). Moreover, for social distancing with $c < 1.5$, a rough estimate of the final fraction of infected molecules is just $\nu_f \sim 2\ (1 - 1/c)$.

5.2. The Height and Width of the Growth-Rate Peak

In the previous section, we found that Eq. (5.6) provides an accurate expression for the final fraction of infected molecules, ν_f, as a function of the parameter, c. Next, we look for analogous expressions for the height and width of the peak in the growth-rate curve.

We begin by referring to Figure 9, which contains plots of the accumulated infection, ν_j, and the infection rate, $\Delta\nu_j$, as a function of the collision step number, j. The left panel

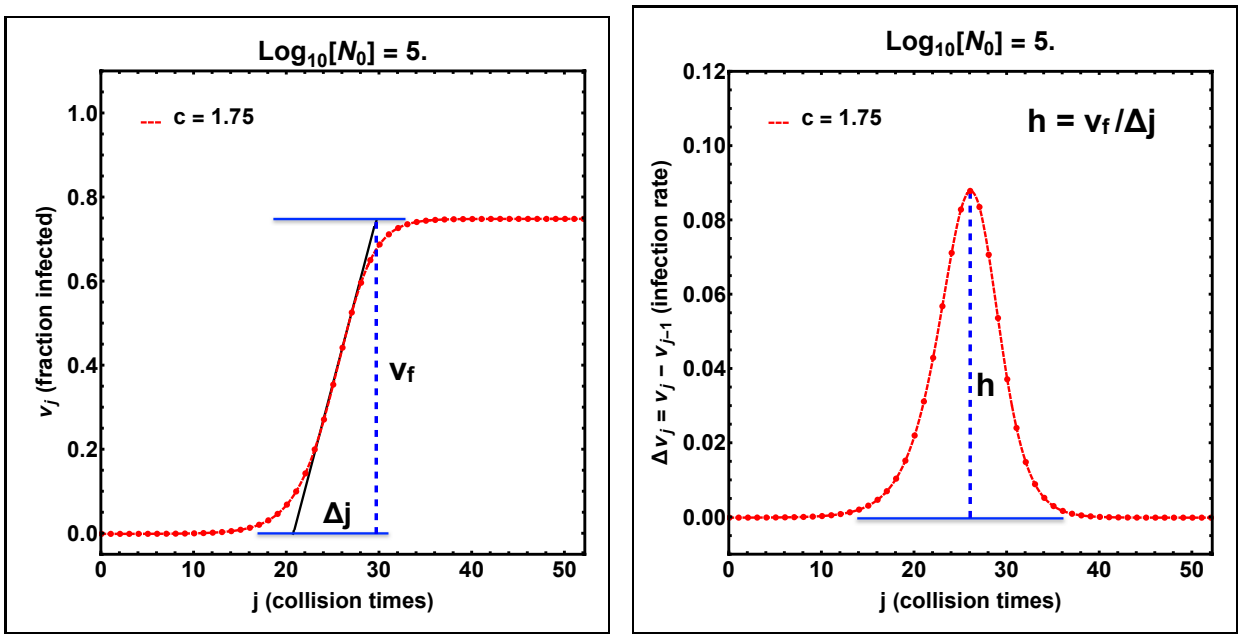

Fig. 9.— Plots of ν_j and $\Delta\nu_j$ for $c = 1.75$, used to illustrate the relation between peak height, $h = \Delta\nu_{max}$, and the width, $w_e = \Delta j$.

shows the accumulated infection curve (in red) and a black line segment drawn tangent to it at the point of maximum slope. Actually, this point is really an extended region, which is why we were able to deduce that $\nu_{j-1} \approx 1 - 1/c$ in the center of this region. (See the discussion leading to Eq. (5.3).) The ends of this line segment lie on the horizontal blue

lines where $\nu_j = \nu_f$ and $\nu_j = \nu_0 = 1/N_0 \approx 0$, and are separated by a horizontal distance, Δj. Consequently, $\nu_f/\Delta j$ is the maximum slope of the ν_j curve, which in turn is equal to the peak value, $\Delta\nu_{max}$, of the growth-rate curve in the right panel. (Here, we ignore the subtle difference between $\Delta\nu_j$ and $d\nu_j/dj$ because $\Delta\nu_j = \nu_j - \nu_{j-1} \approx (\nu_j - \nu_{j-\epsilon})/\epsilon \to d\nu_j/dj$ as $\epsilon \to 0$). Thus, we can write

$$h \;=\; \frac{\nu_f}{\Delta j}, \tag{5.7}$$

where h is the peak height, $\Delta\nu_{max}$.

Next, in Figure 10, we plot the numerical values of the ratio h/ν_f versus the parameter, c, and fit the resulting points with two slightly different formulas given by

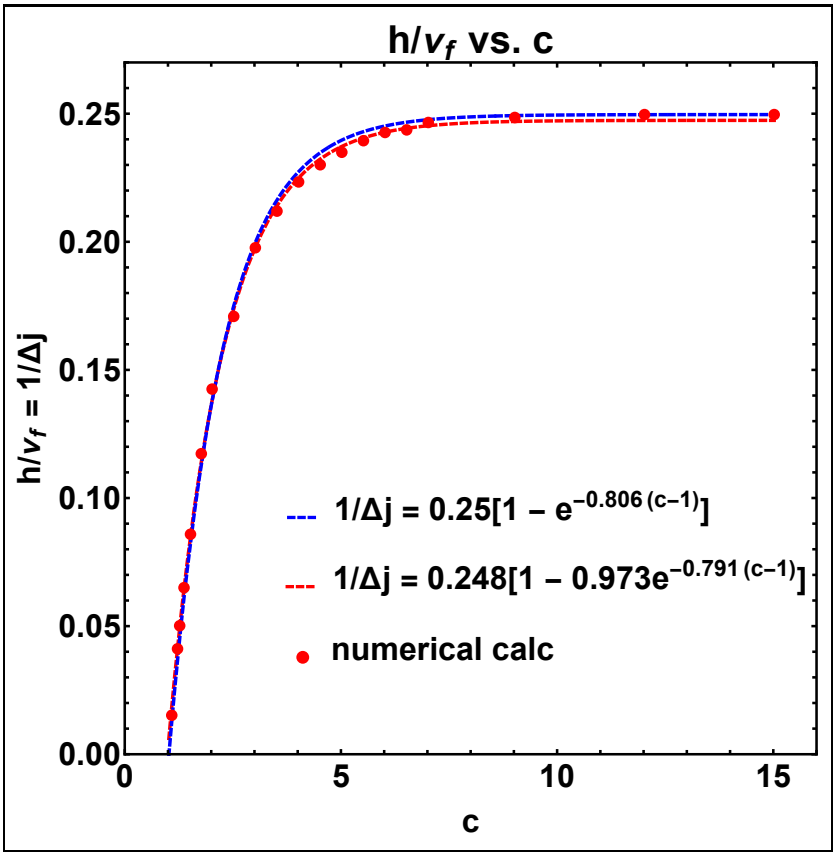

Fig. 10.— Plots of the numerical values of h/ν_f versus c (red points) fit by two different expressions given by Eqs.(5.8a) and (5.8b) (blue and red dashed lines).

$$\frac{h}{\nu_f} \;=\; \frac{1}{\Delta j} \;=\; 0.25\left[1 - e^{-0.806(c-1)}\right], \tag{5.8a}$$

$$\frac{h}{\nu_f} \;=\; \frac{1}{\Delta j} \;=\; 0.248\left[1 - 0.973e^{-0.791(c-1)}\right]. \tag{5.8b}$$

Because Δj is a measure of the width of the growth-rate curve, we refer to it as the width, w_e, of that curve. Consequently, that width can be obtained by taking the reciprocal of the expressions in Eqs.(5.8a) and (5.8b)[1]. As we can see, the simpler formula given by Eq.(5.8a) ends up at 0.25, but misses some of the points at the bend of the curve. However, the more complicated formula given by Eq.(5.8b) passes through the points at the bend of the curve, but ends up slightly lower than 0.25. Both fits pass through all of the points where $c \leq 3$.

The next step is to see how well these expressions combine with Eq.(5.6) for ν_f to fit the calculated points for h. The result is shown in Figure 11. Here, the product of the two fits reproduces the initial curvature in the plot of h versus c. Also, as expected, the better

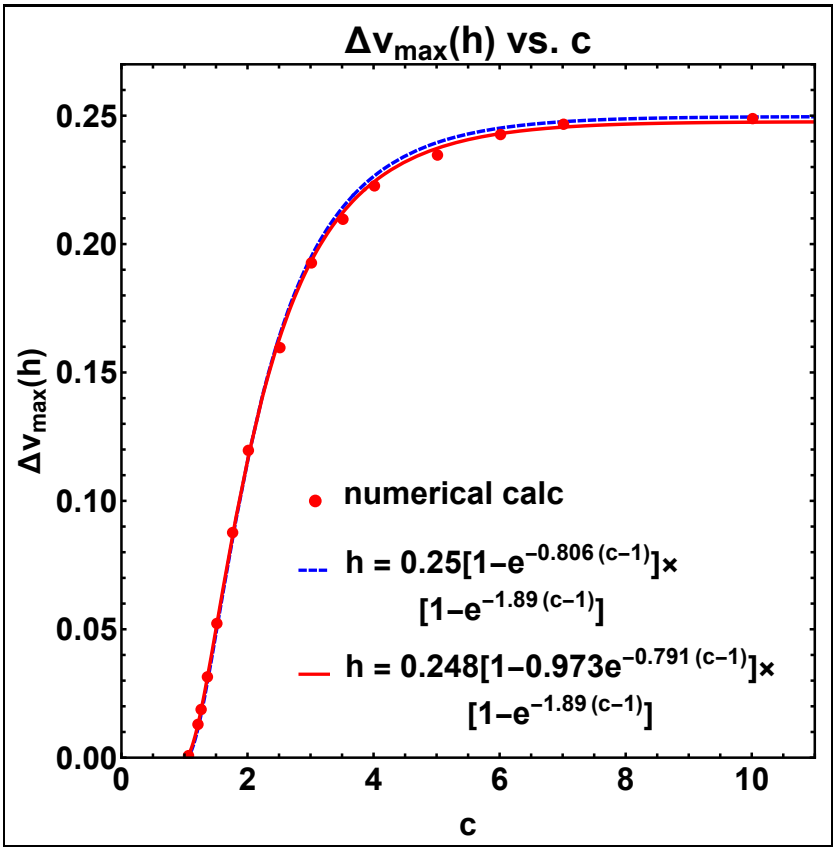

Fig. 11.— Calculated values of $\Delta\nu_{max} = h$, plotted versus c (red points), fit by the product of Eqs. (5.6) and (5.8a) (dashed blue line) and (5.8b) (solid red line).

[1]Other measures of the width like the full-width at half maximum are slightly smaller than w_e. For example, the width, w, would be given by $w = \ln(1+\sqrt{2})w_e = 0.881w_e$ if the fraction of infected molecules, ν_j, were fit by a hyperbolic tangent of the form $\nu_j = (\nu_f/2)[1+\tanh\{\gamma(j-j_{max})\}]$, and the width would be $w = (2^{2/3}-1)^{1/2}w_e = 0.766w_e$ if ν_j were fit by a square-root expression of the form $\nu_j = (\nu_f/2)[1+\gamma(j-j_{max})\{1+\gamma^2(j-j_{max})^2\}^{-1/2}]$, where $\gamma = 2/w_e$.

fit to $1/\Delta j$ in Figure 10 provides the better fit to h in Figure 11. However, even the simpler expressions given by Eqs.(5.6) and (5.8a) provide an essentially perfect fit for $c \leq 3$, which is the region of interest when a substantial amount of social distancing is maintained.

In summary, we can describe the growth curves of Figure 6 and growth-rate curves of Figure 7 in terms of the contagious lifetime parameter, c, using the approximate relations for ν_f, $1/w_e$, and h:

$$\nu_f \;=\; 1 \;-\; e^{-1.890(c-1)}, \tag{5.9a}$$

$$\frac{1}{w_e} \;=\; 0.25\left[1 - e^{-0.806(c-1)}\right], \tag{5.9b}$$

$$h \;=\; \frac{\nu_f}{w_e} \;=\; 0.25\left[1 - e^{-0.806(c-1)}\right]\left[1 - e^{-1.890(c-1)}\right]. \tag{5.9c}$$

5.3. Location of the Peak of the Growth Rate

I have used plots like those shown in Figure 7 to measure the locations, j_{max}, of the peaks for values of c ranging from 1.25 to 10.0, and values of $\log_{10} N_0$ ranging from 2 to 8. Figure 12 shows the measured values of j_{max} plotted versus $\log_{10} N_0$ for given values of c. The dashed lines are the least-squares best fits to these points given by:

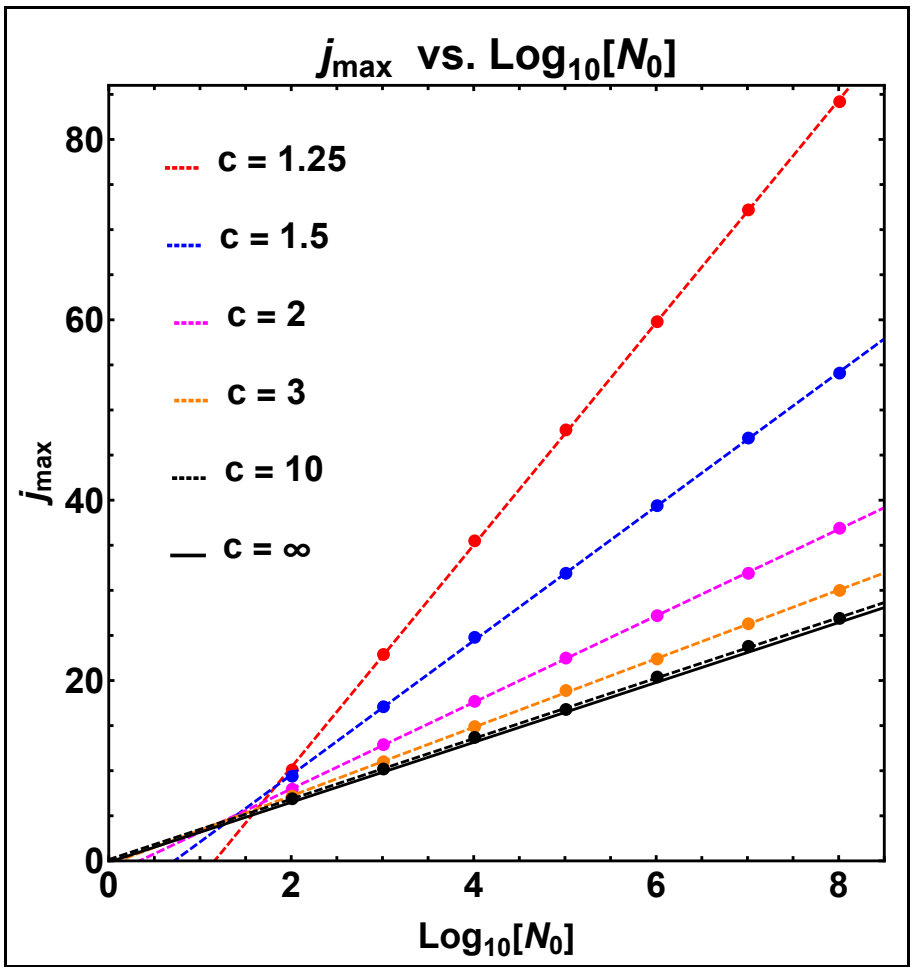

Fig. 12.— Measured j_{max} values plotted versus $\log_{10} N_0$, showing the linear relation between these two quantities for values of $c \geq 1.25$. The line marked $c = \infty$ corresponds to Eq. (3.6).

$$j_{max} \;=\; \begin{cases} -14.071 \;+\; 12.357\;\log_{10} N_0, & c = 1.25, \\ -5.154 \;+\; 7.439\;\log_{10} N_0, & c = 1.5, \\ -1.381 \;+\; 4.799\;\log_{10} N_0, & c = 2, \\ -0.156 \;+\; 3.796\;\log_{10} N_0, & c = 3, \\ +0.312 \;+\; 3.359\;\log_{10} N_0, & c = 10. \end{cases} \tag{5.10}$$

Also, the black line represents Eq.(3.6) ($j_{max} = \log_2 N_0 \approx 3.322 \log_{10} N_0$) for permanently contagious molecules with $c = \infty$. Its slope (3.322) and intercept (0.000) are very close to the slope (3.359) and intercept (0.312) of the line with $c = 10$.

Based on this figure, there can be little doubt that $j_{max} = p_c + q_c \log_{10} N_0$ for each value of the parameter, c, where p_c and q_c are the least squares best-fit parameters for the dashed lines that pass through these points. However, an even more interesting property of Figure 12 is that the spread of the j_{max} values increases with $\log_{10} N_0$, and that this spread tends to a very small minimum value near the point $(1.5, 6)$ where the straight lines nearly intersect. This means that for a small population, the maximum growth rates occur at approximately the same time regardless of the value of c and the corresponding amount of social distancing. This contrasts with the result for a large population for which the location of the peaks depends much more strongly on the amount of social distancing. Thus, to achieve the same maximum growth rate, the larger population must wait a much longer time than a smaller population. As we shall see in subsection (5.3.2), this property can be used to deduce j_{max} as a function of $\log_{10} N_0$ and c. But first in subsection (5.3.1), we will deduce an expression for j_{max} using the linearized solution of the growth equation given in Appendix A.

5.3.1. The slope-intercept method

Appendix A provides a way to estimate the values of p_c and q_c and thereby reproduce the j_{max} expressions in Eq. (5.10). Referring to Figure 25 in Appendix A, we see that the linearized solutions give the initial behavior of ν_j for each value of c. Because these linearized solutions lack the damping effect of the non-linear term, they eventually diverge from the non-linear solutions whose curves bend over and approach final values of ν_f. However, the plots of the linearized solutions in Figure 25 seem to cross the $\nu_j = \nu_f$ threshold approximately where the non-linear solutions have their maximum slopes. Such a relation would provide a way of determining the location of those maximum slopes and therefore the locations of the peaks of the corresponding growth-rate curves of $\Delta\nu_j$. For this purpose, we would simply equate the linearized solution, $\sigma_c \rho_c^j \nu_0$, and the final value of the growth curve, ν_f, and then

solve for j. The first step gives

$$\sigma_c \, {\rho_c}^j \;=\; \frac{\nu_f}{\nu_0} \;=\; \nu_f N_0, \tag{5.11}$$

where ν_f is obtained from Eq.(5.6), and the parameters σ_c and ρ_c can be obtained from Table 1 of Appendix A (or from Eq. (A9) for values of c in the range 1-2), or from Eqs. (C1a) and (C1b) of appendix C. The resulting value of j is $j_{max} - 1$ because this analysis is based on the growth curves in Figure 6 and Figure 25, whereas the peak value of the growth-rate curve comes from the difference, $\Delta\nu_j = \nu_j - \nu_{j-1}$ shown in Figure 7. However, because this method is based on a rough estimate, I would be surprised if the accuracy of the resulting value of j_{max} were closer than 1 unit of collision time. This is comparable to the differences between $\Delta\nu_j$ and $d\nu_j/dj$, which Figure 7 shows are often about 0.5 of a unit.

Solving Eq. (5.11) for j and setting $j_{max} = j + 1$, we obtain

$$j_{max} \;=\; \left\{1 + \frac{\log_{10}(\nu_f/\sigma_c)}{\log_{10}(\rho_c)}\right\} \;+\; \left\{\frac{1}{\log_{10}(\rho_c)}\right\} \log_{10} N_0, \tag{5.12}$$

which is in the format $p_c + q_c \log_{10} N_0$ used in Eq. (5.10). Although, I have used logarithms to base 10 for comparison with the expressions in Eq. (5.10), because these logarithms occur as ratios, they can be expressed in any base, including base e, base 2, or even base ρ_c, without affecting the value of j_{max}.

Next, obtaining ν_f from Eq. (5.6) and substituting the values of σ_c and ρ_c from Table 1 of Appendix A, we find that

$$j_{max} \;=\; \begin{cases} -14.934 \;+\; 12.239 \;\log_{10} N_0, & c = 1.25, \\ \\ \;\;-4.802 \;+\; 7.383 \;\log_{10} N_0, & c = 1.5, \\ \\ \;\;-0.668 \;+\; 4.785 \;\log_{10} N_0, & c = 2, \\ \\ \;\;+0.463 \;+\; 3.780 \;\log_{10} N_0, & c = 3, \\ \\ \;\;+1.000 \;+\; 3.322 \;\log_{10} N_0, & c = \infty. \end{cases} \tag{5.13}$$

These values are comparable to the least-squares best-fit values in Eq. (5.10). The coefficients of $\log_{10} N_0$ are essentially the same. However, the constant terms in Eq. (5.13) are systematically a little larger than those in Eq. (5.10). For the larger values of c, this discrepancy is probably due to the intrinsic differences between the digital peak and the peak of the continuous curve. For the smaller values of c, and in particular for $c = 1.25$, the discrepancy may reflect the weaker and broader profiles whose peak locations are less well defined.

Figure 13 shows the j_{max} values calculated using the full non-linear equation (Eq. (4.4)) and the values obtained with the approximate method given by Eq. (5.12), both using the nominal value of $\log_{10} N_0 = 5$. The red dots are based on Eq. (4.4), whereas the blue dots and their interpolated curve are based on Eq. (5.12). The red and blue points all lie on the same dashed blue curve, indicating a clear dependence of j_{max} on c. For $c \geq 1.4$, the corresponding red and blue points of each pair are very close together, often occulting each other with virtually identical values. However, for $c < 1.4$, the curve steepens and the separations systematically increase as c approaches 1.

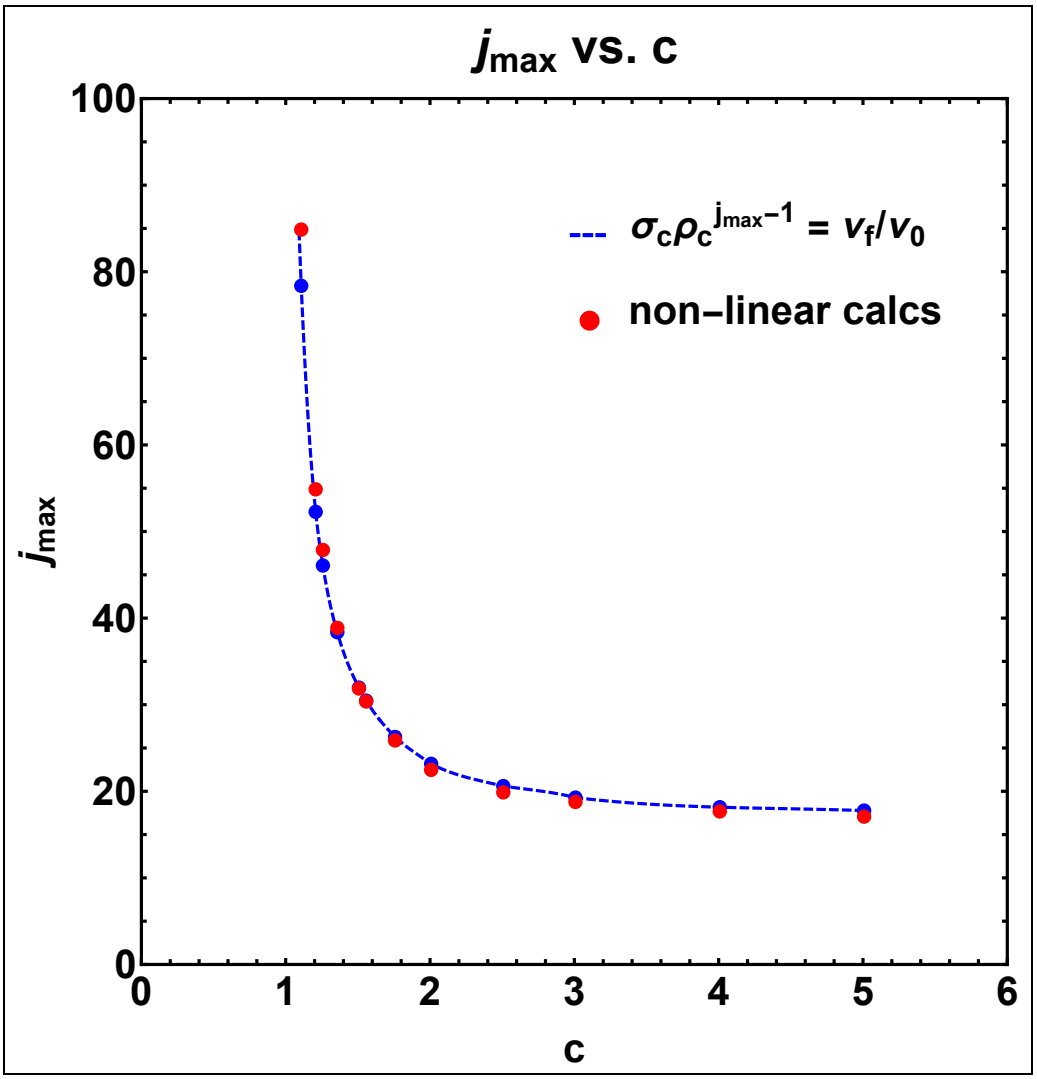

Fig. 13.— j_{max} values, obtained from the non-linear Eq. (4.4) (red points) and calculated with Eq. (5.12) (blue points and dashed line), plotted versus c for $\log_{10} N_0 = 5$.

At first, one might suppose that the overall agreement between the calculated values in Eq. (5.13) and the measured values in Eq.(5.10) supports our ansatz that the point of maximum slope occurs where the linearized solution for ν_j equals the final value ν_f. However,

the agreement was better for q_c than for p_c, which contains the entire dependence on ν_f. So the discrepancy in the values of p_c probably reflects the eventual breakdown of this assumption, especially as c approaches 1 and the growth-rate curves become increasingly flat. Next, we will abandon this assumption, and see if the nearly common intersection of the lines in Figure 12 gives a point-slope formula for j_{max} that is more accurate than the slope-intercept approach.

5.3.2. The point-slope method

In this approach, we return to Figure 12 where the measured values of j_{max} give linear fits that nearly intersect at a common point near $(x_0, y_0) = (1.5, 6)$. We begin by precisely determining the point of closest approach to the lines with $c = 1.25$, 1.5, 2, 3, and ∞, leaving out $c = 10$ because it is essentially the same as $c = \infty$.

We do this in the following way. We recognize that even though these five lines may not intersect in the same point, lines drawn parallel to them can be adjusted so that they do intersect at a common point. This can be done for many such common locations and our objective is to find the one that makes the root-mean-square distance between all those parallel lines a minimum. (Because one set of lines passes through a common point, (x_0, y_0), this is equivalent to making the mean-squared distance from that point to the original lines a minimum.) Consequently, we derive an expression for the mean squared distance, $D^2(x_0, y_0)$, between all of the parallel lines. Then we minimize $D^2(x_0, y_0)$ with respect to x_0 and y_0, and obtain two linear equations for x_0 and y_0. When we solve those equations simultaneously, we obtain $x_0 = 1.586$ and $y_0 = 6.153$. Using these values to evaluate $D^2(x_0, y_0)$ and taking the square root, we find that $D_{rms} = 0.054$.

Next, we write the equation for the lines that pass through the point (x_0, y_0) with the same slope, $q_c = 1/\log_{10}\rho_c$, that we used in the previous subsection. This point-slope approach gives the result

$$j_{max} \;=\; y_0 \;+ q_c(\log_{10} N_0 - x_0) \;=\; (y_0 - x_0 q_c) \;+\; q_c \log_{10} N_0, \tag{5.14}$$

where $(x_0, y_0) = (1.586, 6.153)$ and $q_c = 1/\log_{10}\rho_c$. The quantity, ρ_c, is given by Table 1 of Appendix A. It is the largest positive root of the equation $r^{c+1} - 2r^c + 1 = 0$ when c is an integer greater than 1. However, when c is not an integer, it is necessary to interpolate the linearized difference equation to obtain an equation for r, as we did in Appendix A for c between 1 and 2. An approximate expression is given in Eq.(C1a) of appendix C.

Figure 14 shows these lines plotted with the measured points from Figure 12. The passage of these lines through the common point, (x_0, y_0), is clearly visible. These lines fit the measured points almost as well as the best-fit lines in Figure 12, verifying that j_{max} is

well represented by a line of slope $1/\log_{10}\rho_c$, passing through a point, (x_0, y_0), that does not depend on c. Eq.(5.15) gives numerical values for comparison with Eqs.(5.13) and (5.10).

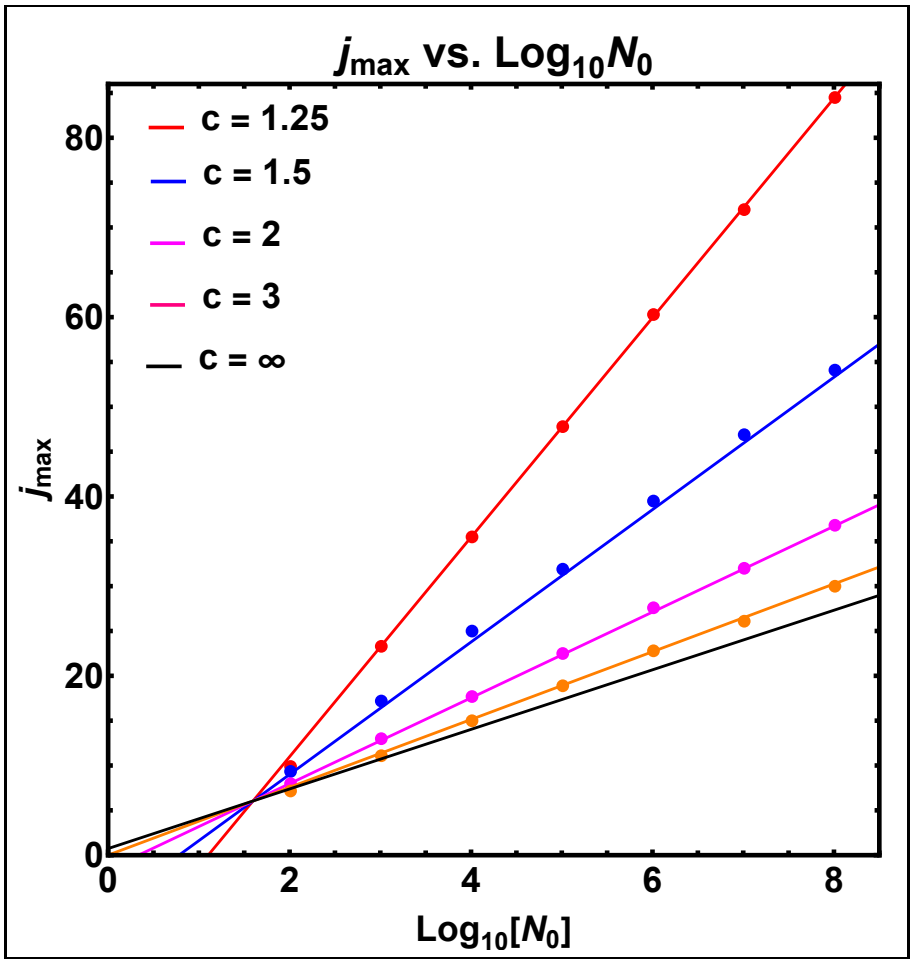

Fig. 14.— j_{max} values calculated using Eq.(5.13) and plotted versus $\log_{10} N_0$ for values of $c \geq 1.25$ for comparison with the straight lines shown in Figure 12.

$$j_{max} = \begin{cases} -13.258 + 12.239 \log_{10} N_0, & c = 1.25, \\ -5.576 + 7.383 \log_{10} N_0, & c = 1.5, \\ -1.436 + 4.785 \log_{10} N_0, & c = 2, \\ +0.158 + 3.780 \log_{10} N_0, & c = 3, \\ +0.884 + 3.322 \log_{10} N_0, & c = \infty. \end{cases} \tag{5.15}$$

6. Red, Green, and Blue Molecules

6.1. Contributions of the red and green molecules

It is interesting to interpret the growth equation in terms of the fractions of red, green, and blue molecules. For $j \leq c$, all of the infected molecules are still contagious (red) and the incremental increase of infections (normalized to the population, N_0) from one collision to the next is

$$\Delta\nu_j \;=\; \nu_j \;-\; \nu_{j-1} \;=\; \overbrace{\nu_{j-1}}^{red}\overbrace{(1-\nu_{j-1})}^{blue}. \tag{6.1}$$

Here, ν_{j-1} is the fraction of infected molecules at time $j-1$ (all of which are red) and $1-\nu_{j-1}$ is the fraction of noninfected blue molecules, which is also the probability of colliding with a blue molecule. Consequently, the incremental increase in infections is just the product, $\nu_{j-1}(1-\nu_{j-1})$. However, for $j \geq c+1$, the molecules that were infected c steps earlier are no longer contagious. Consequently, at step $j-1$, the fraction of infected, but noncontagious, green molecules is ν_{j-1-c}, and the incremental increase becomes

$$\Delta\nu_j \;=\; \overbrace{(\nu_{j-1}-\underbrace{\nu_{j-1-c}}_{green})}^{red}\overbrace{(1-\nu_{j-1})}^{blue}. \tag{6.2}$$

Thus, after a time, c, the increase in the number of infections is still equal to the product of the numbers of red and blue molecules, but this increase is smaller than it would have been if the red molecules were permanently contagious and there were no green molecules. Eq.(6.2) indicates that the increase is smaller because the number of red molecules is reduced by the current number of green molecules.

It is also interesting to consider what happens toward the end of the epidemic when $\Delta\nu_j \to 0$. For permanently contagious molecules, Eq.(6.1) tells us that the factor $(1-\nu_{j-1}) \to 0$ and therefore the other factor $\nu_{j-1} \to 1$. In this case, the epidemic ends because all the blue molecules become infected, and now exist as contagious red molecules. However, if the infected molecules remain contagious for a limited time, $c < 10$, then all of the blue molecules do not become infected and the factor $(1-\nu_{j-1})$ does not approach 0. Instead, the epidemic ends when other factor $(\nu_{j-1}-\nu_{j-1-c}) \to 0$. This means that all of the infected molecules become green and there are no red molecules left to infect blue ones. As we have seen in Figures 6 and 8 and Eq.(5.6), by decreasing c (and thereby increasing the amount of social distancing), we cause a greater number of blue molecules to avoid infection by the time that the epidemic is over.

Figure 15 shows the separate evolutions of red $(\nu_j - \nu_{j-c})$ and green (ν_{j-c}) molecules as well as their sum, ν_j (the combined total of all infections, whether still contagious or not). (We do not need to provide a separate plot for blue molecules because their number, $1-\nu_j$,

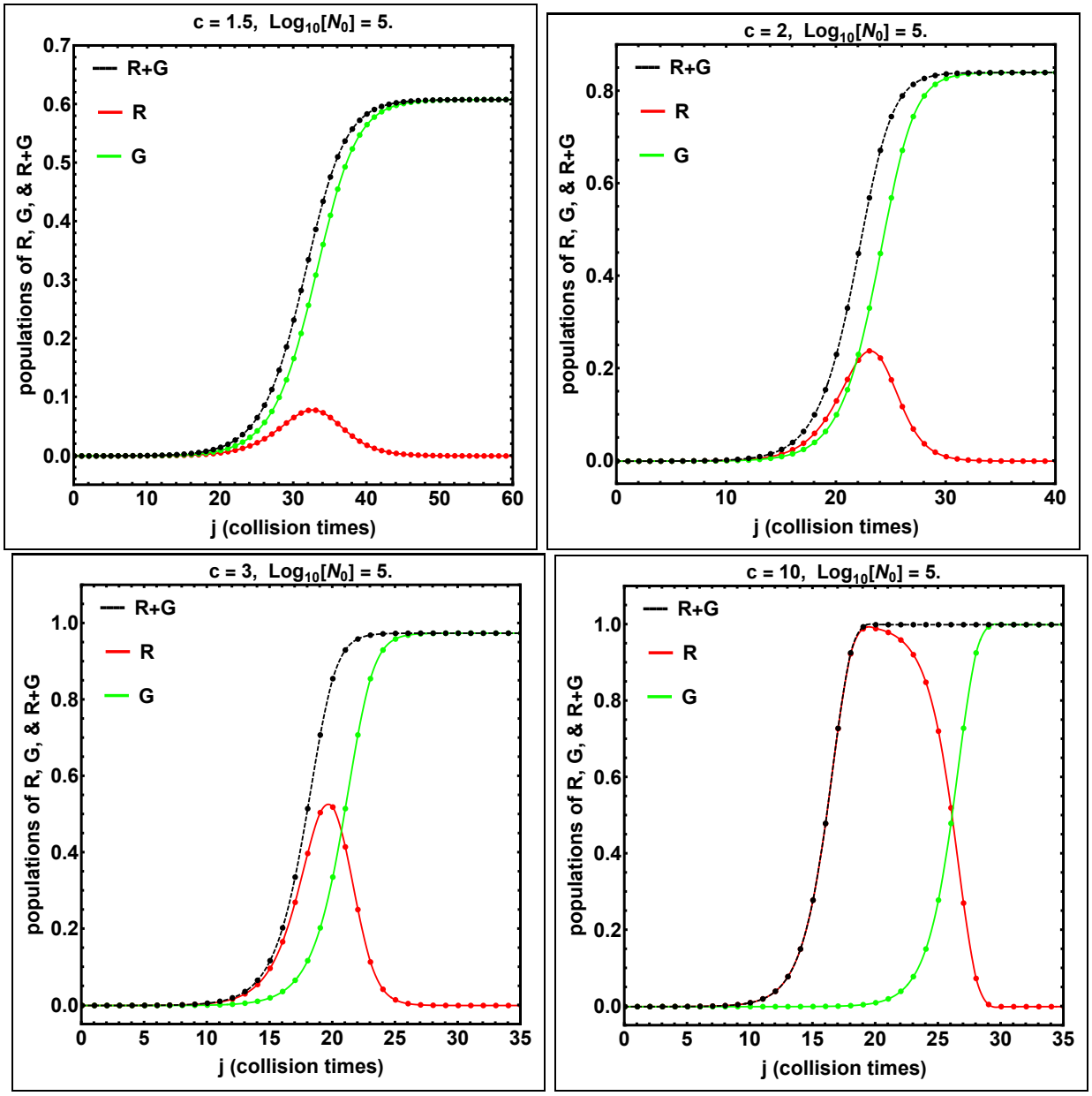

Fig. 15.— Plots of ν_j (Red+Green), $\nu_j - \nu_{j-c}$ (Red), and ν_{j-c} (Green), showing their separate evolutions for $c = 1.5$ (upper left), $c = 2$ (upper right), $c = 3$ (lower left), and $c = 10$ (lower right). Note the scale changes between some of the images.

is just 1 minus the the combined number of red and green molecules, ν_j.) The dashed black curves show the familiar 3-part variation of ν_j as a function of time in units of the step time, τ_s. The change begins slowly for a while, then increases rapidly for a time on the order of $w_e = \Delta j$, and then levels out at a final fraction of infected molecules, whose value, ν_f, depends on the value of c. Because the number of green molecules is given by the lagged quantity, ν_{j-c}, it shows the same time dependence as the number of infected molecules, ν_j, but delayed by c collision times.

The red molecules do not have this 3-part monotonic increase. Shown by the red curves in Figure 15, the number of red molecules have a peaked profile, increasing to a maximum value and then falling back to zero at the end of the epidemic. This is what one would expect from their definition, $\nu_j - \nu_{j-c}$, which is the difference between two terms that eventually become equal.

We can learn more about the time variation of the red molecules by examining the plots in Figure 15 as a function of c. When c is much smaller than the width of the peak (which is essentially the rise time, $w_e = \Delta j$, of the black profile of total infections), we see that

$$\nu_j - \nu_{j-c} \;=\; c(\nu_j - \nu_{j-1}) \;=\; c\Delta\nu_j. \tag{6.3}$$

Thus, for $c << w_e$, the plots of the number of red molecules are within a factor of c of matching the plots of infection rate, $\Delta\nu_j$, in Figure 7. This near equality is shown in Figure 16, which compares $\nu_j - \nu_{j-c}$ and $c\Delta\nu_j$ for $c = 1.5$ and 2 and $\log_{10} N_0 = 5$. In the left panel, the two curves for $c = 1.5$ are nearly identical. In the right panel where $c = 2$, the peak of the red curve is still nearly equal to the height of the black curve, but is now shifted to the right by about 0.5 of a collision time. However, the agreement breaks down as c becomes larger and the steep increase in the number of green molecules is appreciably delayed. For $c \geq 3$, we return to Figure 15 and see that the rising segment of the red curve coincides with the rising segment of the black curve, until the infection rate is dominated by green molecules and the red curve reaches its peak. This leads to the concept of 'herd immunity' which we shall discuss next.

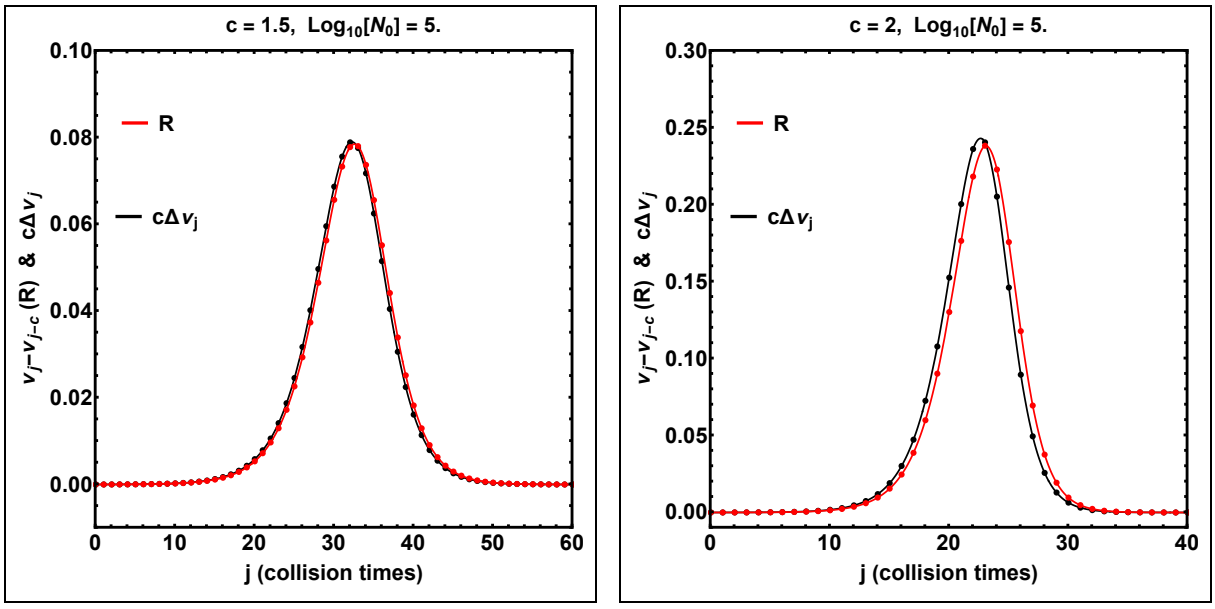

Fig. 16.— Plots of $\nu_j - \nu_{j-c}$ and $c\Delta\nu_j$, comparing the fraction of red molecules with c times the growth rate of infections as c increases from 1.5 to 2.

6.2. Herd 'Immunity'

An equivalent way of thinking about the relation between the numbers of red and green molecules is through the so-called 'herd immunity' concept that one often encounters in the literature (see, for example, Fine et al. (2011)). Here, we think of the green molecules as shielding blue molecules from the remaining red ones, so that some of these red molecules cannot find blue molecules before they lose their contagiousness and turn green. The increased number of green molecules makes it even more difficult for the surviving red molecules to find blue ones, and the number of red molecules starts to decrease. We take the threshold for this turnaround to be the number of infected molecules (relative to the total population, N_0) when the number of red molecules has reached its peak. However, as Eq.(6.2) indicates, the remaining blue molecules are still subject to infection as long as there are red molecules in the population, and the epidemic is not over until all of those red molecules have turned green. Consequently, even after reaching herd immunity, an appreciable time remains before the red population is reduced to 0 and a blue molecule is safe from infection.

We can think about this process graphically with the help of Figure 15. The number of red molecules is given by $\nu_j - \nu_{j-c}$, the difference between the black and green curves. This difference is maximum where $\Delta(\nu_j - \nu_{j-c}) = 0$, which means that $\Delta\nu_j = \Delta\nu_{j-c}$. Consequently, the red curve reaches its peak where the profiles of $\Delta\nu_j$ and $\Delta\nu_{j-c}$ intersect, which is somewhere between the peaks of those two profiles, depending on the value of c. Another way of interpreting the equality, $\Delta\nu_j = \Delta\nu_{j-c}$, is to say that the red curve reaches its peak at the value of j where the slopes of the black and green curves are equal. When c is small, these points lie slightly above and below the points of maximum slope on the black and green curves, corresponding to a relatively small peak in the red curve. As c increases, the peak height of the red curve increases, and the points of equal slope on the black and green curves move farther apart, so that ν_j increases and lies closer to the 'knee' of the black curve and ν_{j-c} decreases and lies closer to the 'toe' of the green curve. Eventually, for $c = 10$, $\nu_j = 1$ and $\nu_{j-c} = 0$, corresponding to a common slope of 0.

To describe this process more quantitatively, we refer to Figure 17, which compares plots of the numbers of red molecules ($\nu_j - \nu_{j-c}$) with corresponding plots of the total number of infections (ν_j) for values of c in the range 1.25 to 20. In the upper panel, we find the time that a particular plot of the number of red molecules reaches its peak, and then in the lower panel, we find the value of ν_j for the corresponding plot of the total number of infections at that time. Thus, in the upper panel, the blue curve with $c = 1.5$ has a peak of strength 0.08 at $j = 33$ which corresponds to the value $\nu_j = 0.38$ in the blue profile in the lower panel. As c increases, the threshold value of ν_j also increases, eventually approaching 1 as c becomes large. In the lower panel, the curves with $c = 5$ and $c = 20$ are almost indistinguishable, indicating that ν_j is already very close to 1. This is consistent with our earlier observation that the growth is essentially the same for $c = 10$ as it is when $c \to \infty$.

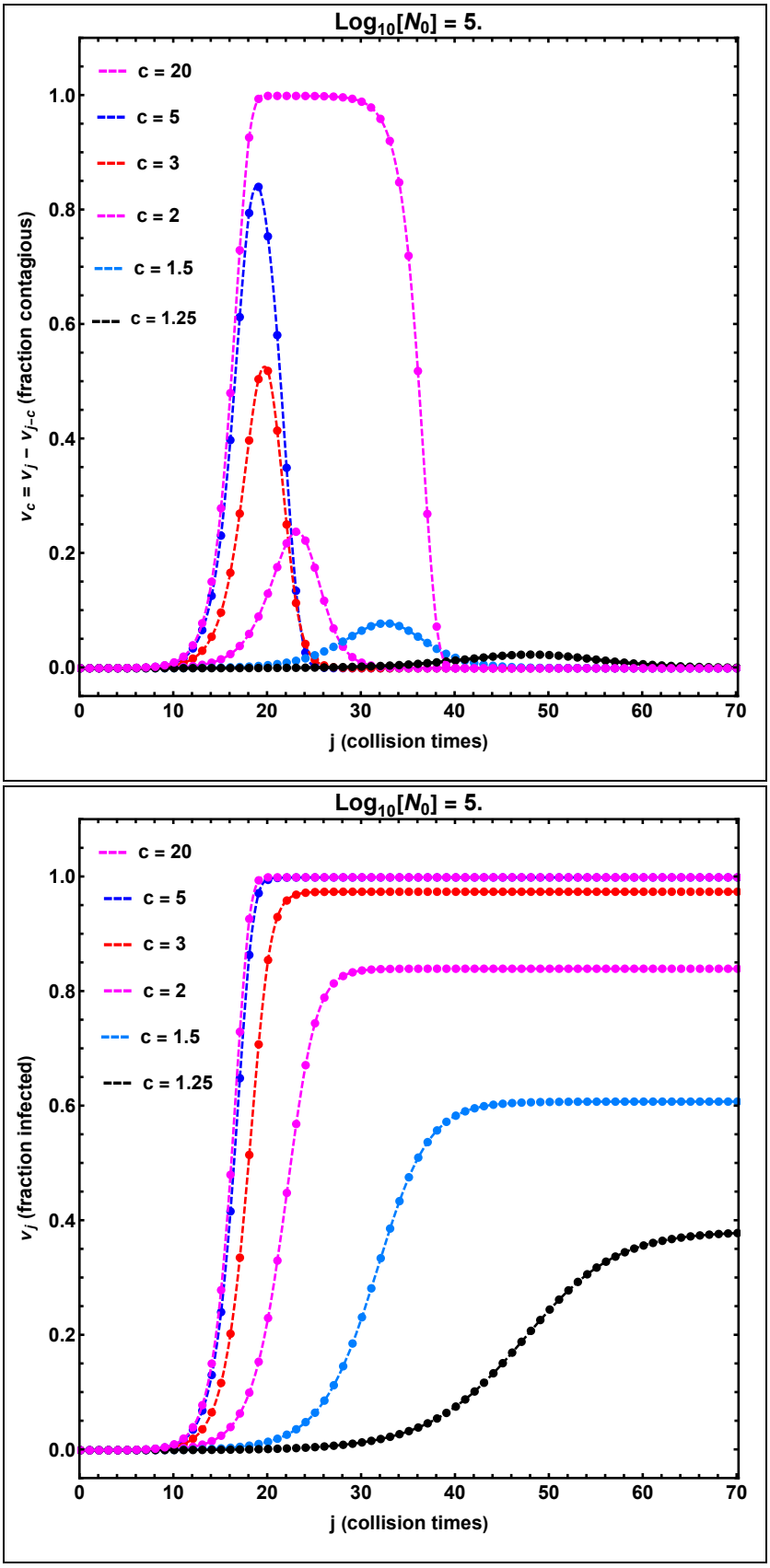

Fig. 17.— The fraction of currently contagious molecules, $\nu_c^{(j)}$ (top), compared with the fraction of infected molecules, ν_j (bottom) in the total population of $N_0 = 10^5$ molecules.

Figure 18 shows a plot of these computed values of ν_j, which we now refer to as the herd-immunity threshold, ν_{herd}, versus the corresponding values of c. The solid red curve is the root-mean-square best fit to those points using the formula

$$\nu_{herd} \;=\; 1 \;-\; e^{-0.860(c-1)}. \tag{6.4}$$

Having learned what the fraction of total infections is when the corresponding fraction of

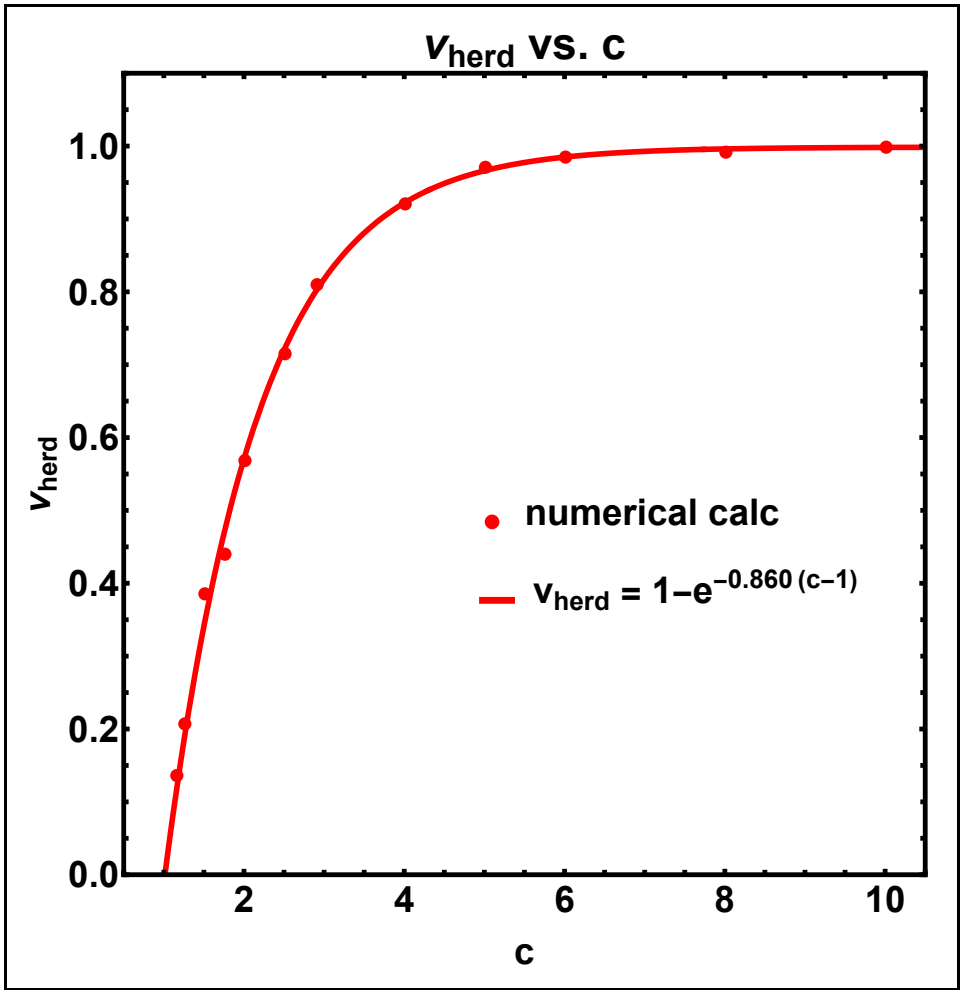

Fig. 18.— A plot of the total number of infections, ν_{herd}, when the fraction of contagious red molecules reaches its peak. The red curve is the root-mean-square best fit to the calculated points (red dots).

contagious red molecules reaches its peak, we now ask when the fraction of red molecules actually reaches its peak. In particular, after reaching the time, j_{max}, that the total number of infections (red plus green molecules) reaches its peak, how much longer must we wait until the number of red molecules reaches its peak? In Figure 16, the plots of $\nu_j - \nu_{j-c}$ and $c\Delta\nu_j$ have already shown that the wait time is roughly 0.5 of a collision time when $c = 2$ and even less when $c = 1.5$. So the next step is to make additional comparisons and plot the results.

Figure 19 shows the result of measuring the peak locations of $\Delta\nu_j$ and $\nu_j - \nu_{j-c}$. The measured lags are relatively small, never exceeding 2 collision times, which is the difference,

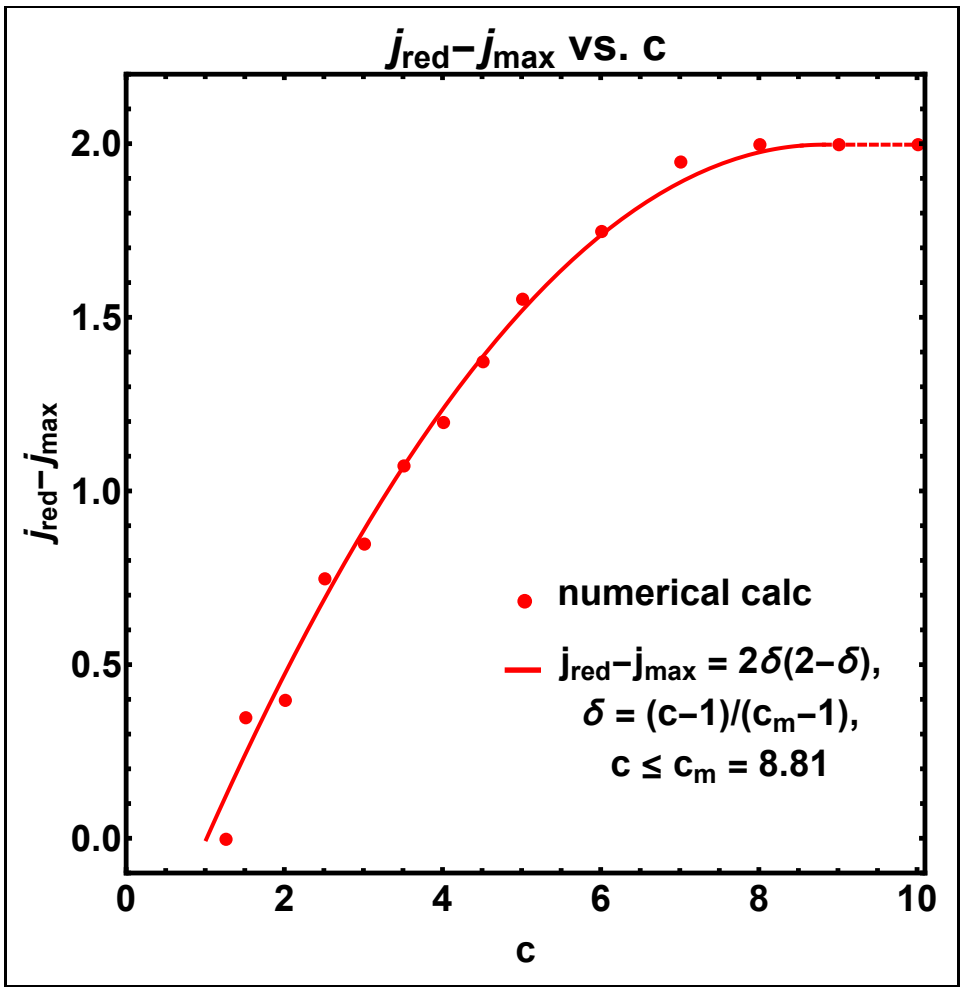

Fig. 19.— The time lag, $j_{red} - j_{max}$, between the peaks of the total number of infections, ν_j, and the number of contagious molecules, $\nu_j - \nu_{j-c}$, plotted versus the contagious time, c. The solid red curve is the root-mean-square best fit to the calculated points (red dots) given by Eq.(6.5), and the dotted red line is its extension with a constant lag of 2 after the solid curve reaches its peak at $c_m = 8.81$.

$j_{red} - j_{max}$, when $c \to \infty$. (For this calculation, we return to Figure 2 and recognize that $j_{red} = 19$ is the j-value when ν_j reaches 1, as shown in the left panel, and that $j_{max} = 17$ is the time that $\Delta\nu_j$ reaches its peak, as shown in the right panel.) Coming back to Figure 19, we note that the data points are best fit by the quadratic function

$$j_{red} - j_{max} \;=\; 2\left(\frac{c-1}{c_m-1}\right)\left\{2 - \left(\frac{c-1}{c_m-1}\right)\right\} \;=\; 0.512(c-1) - 0.033(c-1)^2, \qquad (6.5)$$

where c_m is the value of c where this quadratic function reaches its maximum value of 2 units of the step time, τ_s. In this case, $c_m = 8.81$. The numerical form of this equation shows that

the lag starts as a linear function whose slope is nearly 1/2, but 'decelerates' very slightly as c becomes large. Taking the derivative of Eq.(6.5), we obtain the slope

$$\frac{d}{dc}(j_{red} - j_{max}) \;=\; \left(\frac{4}{c_m - 1}\right)\left(\frac{c_m - c}{c_m - 1}\right) \;=\; 0.512\left(\frac{c_m - c}{c_m - 1}\right), \tag{6.6}$$

which equals 0.51 when $c = 1$ and becomes 0 when $c = c_m$. Thus, if $c - 1$ is small, then $j_{red} - j_{max} \approx 0.51(c-1)$, which gives lags of 0.51 for $c = 2$ and 0.25 for $c = 1.5$, consistent with our measurements from Figure 16.

It is interesting to compare the c-dependence of this lag with the corresponding formula obtained for profiles like the hyperbolic tangent and square-root functions mentioned in the footnote of section 5.2. For example, we can use the hyperbolic tangent formula, $\nu_j = (\nu_f/2)[1+\tanh\{\gamma(j-j_{max'})\}]$, to construct the fraction of red molecules, $R_j = \nu_j - \nu_{j-c}$, and then determine the location of its peak value by setting $\Delta R_j = R_j - R_{j-1} = 0$. The result is $j - j_{max'} = (c+1)/2$, where $j_{max'}$ is the j-value for which ν_j has its greatest slope. j_{max} (without the prime) is where $\Delta\nu_j$ has its peak and differs from $j_{max'}$ by 1 unit according to $j_{max} = j_{max'} + 1$. Consequently, $j_{red} - j_{max} = (c+1)/2 - 1 = (c-1)/2$, which is essentially the same as the starting value of $0.51(c-1)$, that we found in Figure 19 for the RGB-profiles[2].

Despite this agreement between the lag obtained for the RGB-model and the lag obtained for the hyperbolic tangent formula, there is an important difference. Namely, as c increases, the *RGB*-lag gradually deviates from its initially linear c-dependence due to the presence of the small quadratic term in Eq.(6.5). As shown in Figure 19, the plot bends over and eventually saturates at 2 units of collision time when c becomes comparable to the width of the ν_j profile. In contrast, the lag for the hyperbolic tangent model remains linear as c increases until the saturation occurs. This distinction reflects different curvatures of the RGB-profile above and below the point that ν_j has its maximum slope, especially as c becomes large. We have already seen this asymmetry in the plots of ν_j and $\Delta\nu_j$ in Figure 2 when the red molecules were permanently contagious, and in Figures 6 and 7 when $c > 2$. The hyperbolic tangent formula does not have this asymmetry and therefore gives a lag that remains linear as c increases until the saturation occurs.

We can understand why this asymmetry causes $j_{red} - j_{max}$ to deviate from its initially linear dependence by referring to plots of $R + G$ (black curve), G (green curve), and R (red curve) in Figure 15, and considering how the black curve changes around j_p, its point of maximum slope. For $j \geq j_p$, the slope can be approximated by $s_j = s_p - a_2(j - j_p)$, where a_2

[2]Note that to obtain this expression, we had to determine the red peak from $\Delta R_j = 0$, rather than $dR_j/dj = 0$, which gives $j = j_{max'} + c/2$, and we had to distinguish between the place, $j_{max'}$, that ν_j has its greatest slope, and the place, $j_{max} = j_{max'} + 1$, that $\Delta\nu_j$ has its peak. These small differences matter when the range of $j_{red} - j_{max}$ is only 2 units of τ_s.

is the curvature (*i.e.* the rate of change of the slope with respect to j) on the upper branch of the profile. Likewise, $s_j = s_p - a_1(j_p - j)$ on the lower branch where $j_p \geq j$. We assume that both a_2 and a_1 are positive so that the slope decreases on both sides of j_p, as it ought to do at a point of maximum slope.

The essential point here is that for each value of j, the red curve is produced by subtracting contributions from the upper branch of the black curve where $j = j_p + \delta$ and the lower branch of the (identical, but shifted) green curve where $j = j_p - c + \delta$. Consequently, if c is small and the black and green curves lie close together, then the peak of the red curve will lie at the mid-point between the black and green curves where $\delta = c/2$ and the slopes are both very close to the peak value s_p. (As noted above, the condition $\Delta R_j = 0$ adds an extra half step, and the reference to j_{max}, rather than j_p, subtracts a full step, so that the result is $j_{red} - j_{max} = (c-1)/2$, rather than $c/2$.)

However, when c becomes large, the slope will be $s_p - a_2(j - j_p - \delta)$ on the upper branch of the black curve and $s_p - a_1(j_p - j - c + \delta)$ on the lower branch of the green curve. To find the value of j for which these slopes are equal, we simply equate those expressions and solve for j, obtaining $j - j_p = \{a_1/(a_1+a_2)\}c$. If the accelerations are equal, then $j = j_p + (c/2)$ (or $j_{red} - j_{max} = (c-1)/2$ for the condition $\Delta R_j = 0$). But if $a_2 > a_1$, then $j - j_p < c/2$, and the location of the red curve falls behind the mid-point by the amount $(c/2)\{(a_2 - a_1)/(a_2 + a_1)\}$. This asymmetry causes the lag to deviate from its initially linear dependence on c that we found in Figure 19.

In summary, the 'herd immunity' occurs when red molecules are turning green as fast as they are being produced by the infection of blue molecules. At this time, the number of red molecules has reached its peak and the infection rate of blue molecules is slightly past its peak. In fact, it is convenient to think in terms of the time that the blue molecules are being removed most rapidly (and the red-plus-green molecules are being created most rapidly). This time occurs approximately half-way up the growth profile for red-plus-green molecules, which means that the red peak will occur somewhere in the time remaining between the middle and end of that profile. As Figure 19 showed, this time lag, $j_{red} - j_{max}$, depends on c, but it is relatively small and always less than 2 collision times, τ_s. Thus, for serious social distancing and a value of c close to 1, the time, j_{max}, will be greatly delayed, as will j_{red}, which follows closely after it. However, the good news is that the immunity threshold, ν_{herd} will be correspondingly small and so will the total number of infected molecules, ν_f, when the epidemic is over.

Finally, it is important to recognize that the 'herd immunity' does not give the blue molecules any immunity against infection by the red ones. It is just a way of saying that the rate of infection has reached its maximum, and will be decreasing for the remainder of the ramp time. Although it may be reassuring to think of the green molecules as shielding blue molecules from red ones, it is not so reassuring to realize that red molecules provide

the shielding when infected molecules are permanently contagious. Once the infection rate has reached its peak, the red molecules are so plentiful that some block others and prevent them from participating in the 'feeding frenzy' that will now continue at a diminishing rate. In the words of former baseball player, Yogi Berra (1998), 'It ain't over 'til it's over' and all the red molecules are gone.

7. Changing Social Distancing During an Epidemic

7.1. Removing the Social Distancing Entirely

An interesting application of these calculations is to see what would happen if the amount of social distancing were relaxed suddenly before the virus has been completely eliminated. Figure 20 shows the result of suddenly changing c from $c = 1.5$ to 10 after 42 collisions when the fraction of infected molecules had increased to within 3.3% of its final value of 0.61, and the infection rate had decreased to about 14% of its peak height. This calculation was done for a population of $N_0 = 10^5$, which means that only $0.033 \times 0.61 \times 10^5 \approx 2000$ more molecules would be infected if the social distancing were to continue at its current rate.

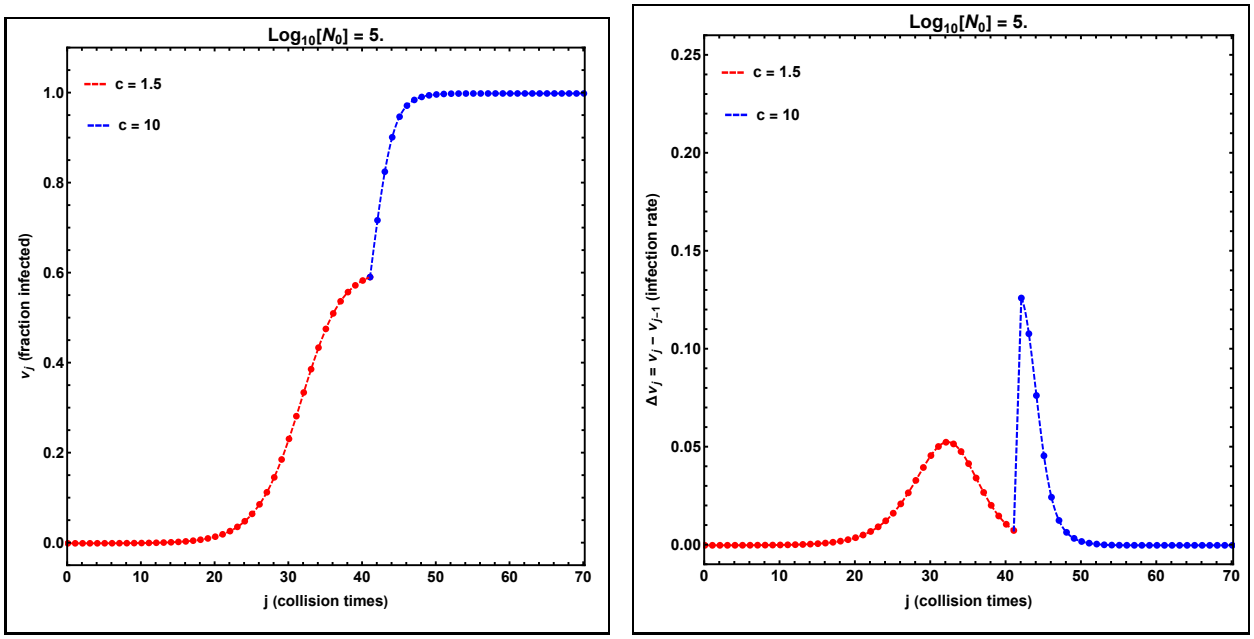

Fig. 20.— The fraction of infected molecules (left), and the infection rate (right) when social distancing is relaxed suddenly from $c = 1.5$ to $c = 10$ after 42 collisions, showing the rapid infection of the remaining blue molecules.

What was the number of contagious molecules at that time? The fraction of contagious molecules is so small that we cannot obtain an accurate estimate from Figure 17. And in that figure, the overlap between the curves with $c = 1.25$ and $c = 1.5$ make the estimate even more difficult. However, we can obtain a fairly accurate estimate from the enlarged plot of ν_j in Figure 6 in combination with Eq.(5.1) for $\Delta\nu_j$. In Figure 6, the slope of the curve with $c = 1.5$ is 0.00657 per step (or collision) length when $j = 40 - 42$. For $c = 1.5$ (corresponding to 1.5 steps), this means that the change $\nu_{42} - \nu_{40.5} = 1.5 \times 0.00657 = 0.00986$. Thus, for a total population of $N_0 = 10^5$ molecules, this corresponds to 986 contagious molecules after $j = 42$ collision times. This number is relatively small compared to the 0.6×10^5 molecules

that had been infected at this time, as we can see in the bottom panel of Figure 17. However, as we will see next, those 986 contagious molecules would soon do great damage.

As shown in the right panel of Figure 20, at $j = 42$, the infection rate increased suddenly from $\Delta\nu_j \approx 0.007$ to $\Delta\nu_j = 0.13$, which is 2.5 times the peak rate that occurred near $j_{max} = 32$. By comparison, the fraction of infected molecules increased rapidly from $\nu_j \approx 0.59$ (near their plateau value of $\nu_j = 0.61$) to $\nu_j = 1$. Thus, even when the infection rate had decreased to less than 1% (which is about 14% of its peak value), the sudden relaxation of social distancing caused a rapid burst of infection that effectively wiped out the 40,000 surviving blue molecules. Clearly, 986 contagious (red) molecules were sufficient to restart the infection at a very high rate (2.5 times the previous peak rate at $j_{max} = 32$).

7.2. Sudden Increase of c

We have seen that $c = 10$ is equivalent to a permanently contagious virus with no social distancing. Now, we consider a presumably more realistic situation in which c changes suddenly from $c = 1.25$ to $c = 2.50$. Figure 21 shows the first case, in which the social

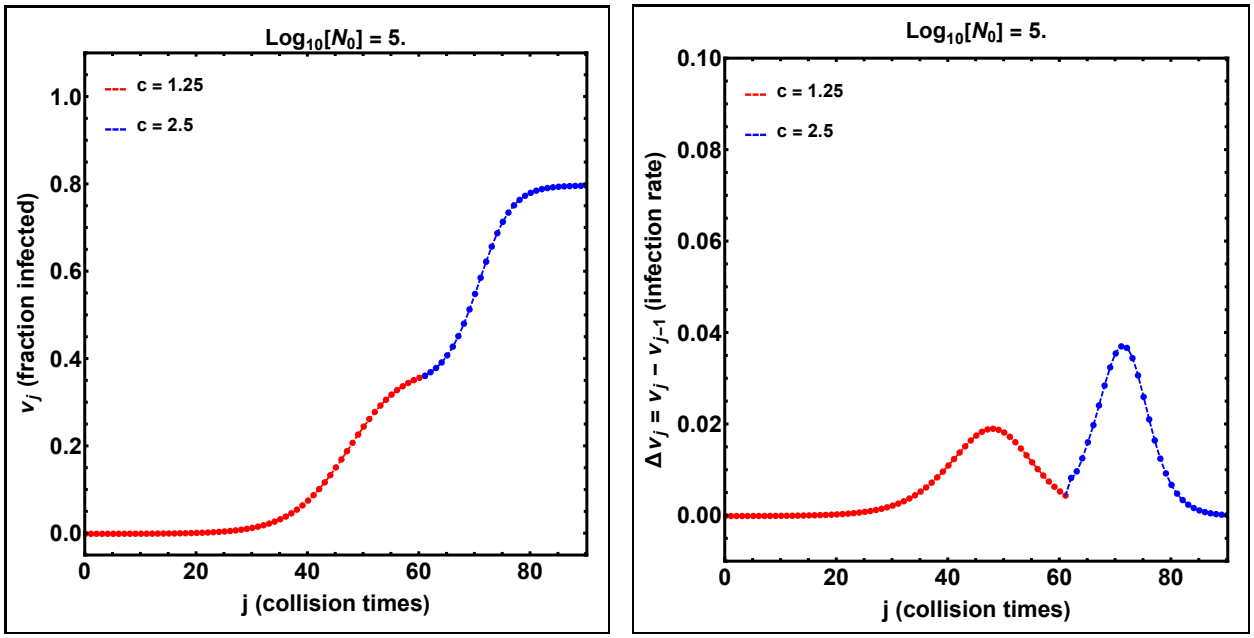

Fig. 21.— The fraction of infected molecules (left), and the infection rate (right) when social distancing is relaxed suddenly from $c = 1.25$ to $c = 2.5$ after 62 collisions when the infection rate was down to 25% of its peak value.

distancing changes after 62 collisions when the fraction of infections is leveling off near $\nu = 0.35$ and the infection rate is down to about 25% of its peak value. In this case, there is a surge in the infection rate, reaching approximately 0.04, which is about 2.0 times

the peak height prior to the change. Also,the fraction of infected molecules changed from 0.35 to 0.80, corresponding to infection of 45,000 extra blue molecules in this population of $N_0 = 1 \times 10^5$. Of course, this increase in the number of infections is proportional to the size of the population and would be higher (or lower) if the population were higher (or lower).

Figure 22 shows the second case, in which c also changes from $c = 1.25$ to $c = 2.50$, but after 48 collisions when the epidemic was near its peak. In this case, there is a much larger surge in the infection rate, reaching approximately 0.11, which is 5.5 times the peak height prior to reducing the amount of social distancing. Also, the fraction of infected molecules reaches about 0.89, corresponding to 54,000 more infections than would have occurred if the social distancing had been maintained at the value of $c = 1.25$. Thus, 9,000 more infections occurred when the change was at $j = 48$ near the peak of the infection rate than when the change was at $j = 62$ toward the end of the epidemic. As before, this enhancement was for a population of $N_0 = 1 \times 10^5$ molecules, and would be proportionally more (or less) for a larger (or smaller) population.

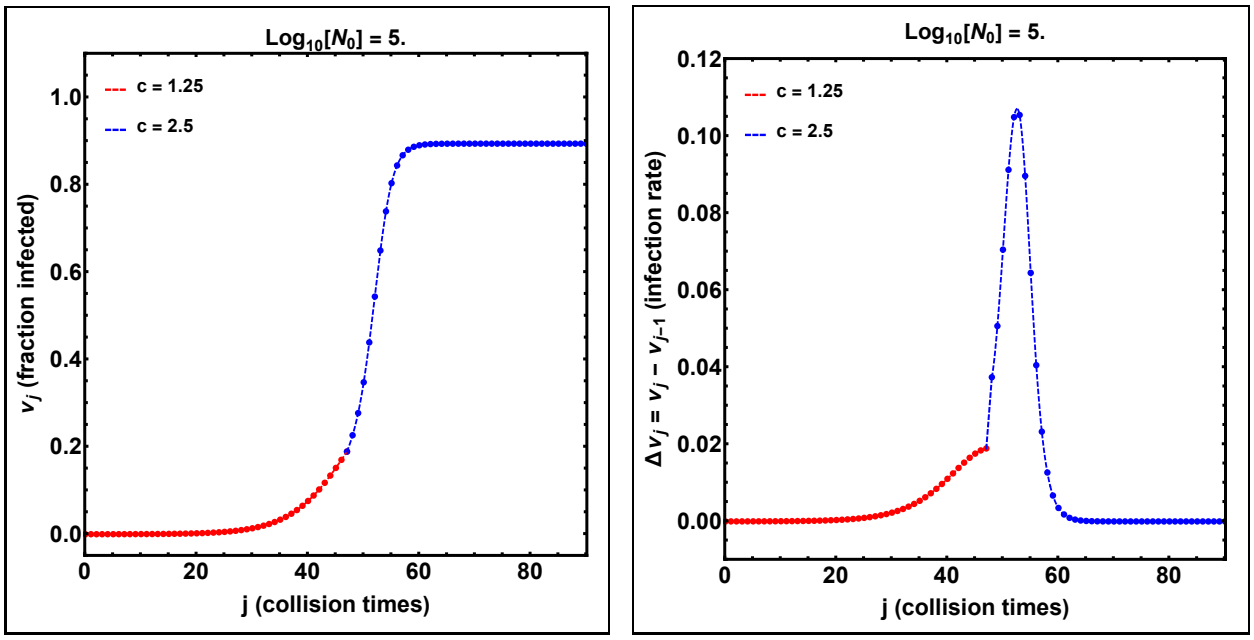

Fig. 22.— The same as Figure 21 except that the social distancing was relaxed after 48 collisions, when the infection rate was near its peak value.

Making the change from $c = 1.25$ to $c = 2.5$ near the peak of the growth rate caused the final fraction of infected molecules to be $\nu_f = 0.89$. This value was larger than the value of $\nu_f = 0.80$ that was obtained in the previous example when the growth rate had fallen to 25% of its peak value. However, $\nu_f = 0.89$ is still less than the final value of $\nu_f = 0.93$ that would be obtained if the epidemic proceeded from the beginning with $c = 2.5$. Equivalently, $\nu_f = 0.93$ would have been obtained if the transition occurred for $j \leq 35$ toward the start of

the epidemic with $c = 1.25$. In fact, the more that the transition is delayed, the smaller the final fraction, ν_f, becomes until the limit of $\nu_f = 0.77$ is reached for $j \approx 75$. (Nevertheless, $\nu_f = 0.77$ is still larger than the value of $\nu_f = 0.37$ that would have been obtained if the change had not been made and the virus had progressed to completion with the original value of $c = 1.25$.) Likewise, the maximum growth rate decreases from $\Delta\nu_j = 0.16$ to $\Delta\nu_j = 0.027$ when the transition time moves from $j = 35$ to $j = 75$.

7.3. Sudden Decrease of c

Now, let us see what happens when the value of c is decreased suddenly from $c = 2.5$ to $c = 1.25$, corresponding to an increase of social distancing. Figure 23 shows the result when the change occurs prior to the maximum in the infection rate while $c = 2.5$. Instead of continuing up to the final value $\nu_f = 0.93$, the fraction of infections levels off at $\nu_f = 0.65$, as shown in the left panel. Also, as shown in the right panel, the infection rate abruptly stops its rise and begins a more rapid decent toward 0.

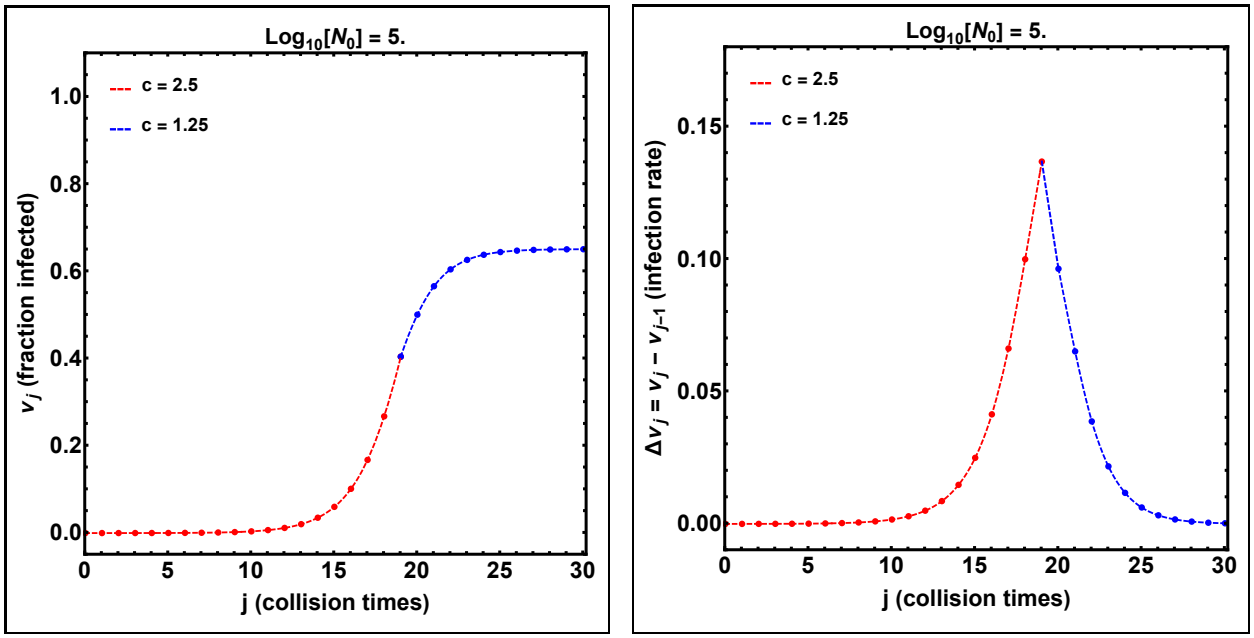

Fig. 23.— Similar to Figure 22 except that the change is done in reverse order with $c = 2.5$ changing to $c = 1.25$ at $j = 20$ before the infection rate reaches its maximum.

Figure 24 shows the effect of making the change at $j = 22$, after the peak in the infection rate. Again, the fraction of infections levels off quickly, but this time at $\nu_f = 0.80$, which is a larger value than we obtained by making the change earlier. Also, the infection rate falls suddenly, but too late to influence the height of the peak. So if we increase the amount of social distancing, we will obtain fewer total infections and a lower peak infection rate if we

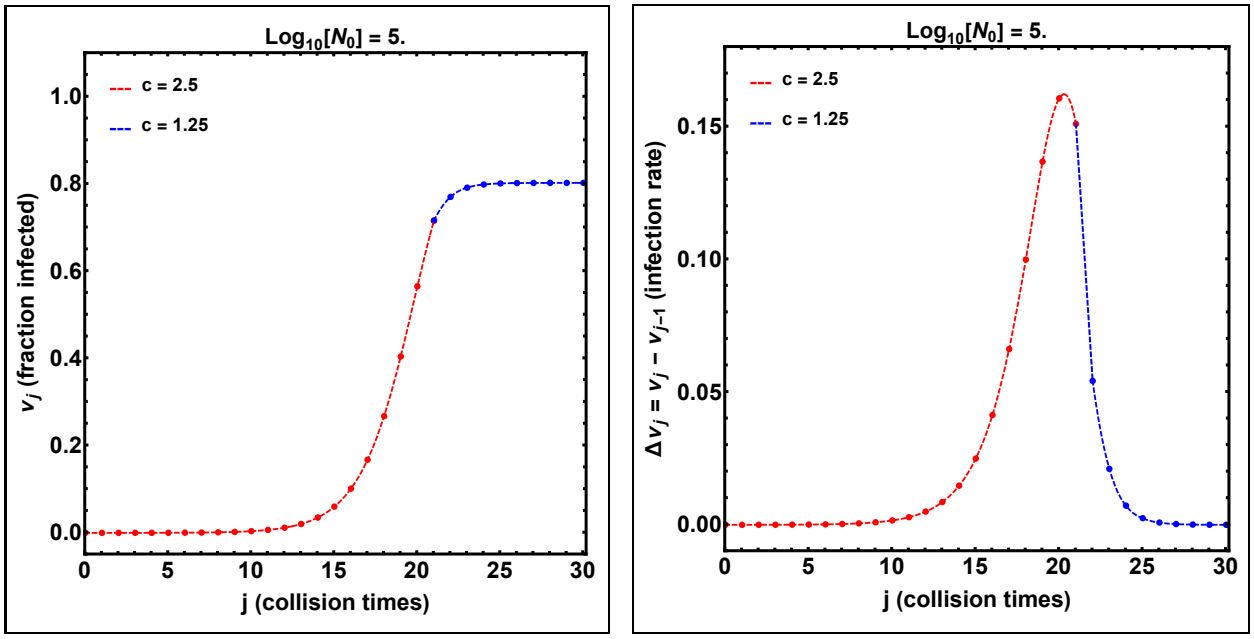

Fig. 24.— Similar to Figure 23 except that the transition occurs at $j = 22$, after the maximum of the infection rate.

make the change early before the infection rate reaches its peak than if we wait until after the peak has occurred. This is opposite of what we found in the previous section where the amount of social distancing was suddenly decreased. In that case, it was necessary to wait until well after the peak to keep the final number of infections and the infection rate low. Thus, to minimize the total number of infections, social distancing should be applied as soon as possible when entering an epidemic and it should be removed as late as possible when coming out of the epidemic.

8. Summary

This paper describes a mathematical model for the spread of a virus through a population represented by the colliding molecules of a gas. In this model, uninfected molecules are colored blue, contagious molecules are red, and molecules that were infected, but have lost their ability to infect are colored green. The epidemic starts with a single red molecule entering a gas of $N_0 - 1$ blue molecules. The red molecule collides with a blue molecule, which becomes infected and turns red. Then the two red molecules collide with two more blue molecules making them red, and the process continues, rapidly increasing the number of infected molecules in the gas.

If infected molecules remained contagious forever, the fraction of infected molecules would continue to increase exponentially until about half of the molecules are infected.

Then, with fewer blue molecules left, the infection rate would decrease and the epidemic would end when all of the remaining blue molecules are infected. No one has escaped. However, if an infected molecule loses its contagiousness after a finite time, τ_c, then the spread of the infection is delayed and weakened, and the epidemic ends without infecting all of the blue molecules. The spreading rates, times, and magnitudes all depend on the ratio of the contagious lifetime, τ_c, to the average time between collisions, τ_s, the step rate used in the calculation. Thus, the key parameter is $c = \tau_c/\tau_s$.

In principle, we could fix τ_s and let τ_c vary, as if we were studying a variety of different diseases. However, I have regarded τ_c to be fixed, and imagined that the collision time, τ_s, varies with c due to different amounts of social distancing for a single disease.

After setting up the model with permanently contagious molecules, I performed the numerical computations and graphed the solutions, which included the evolution of the number of infected molecules as well as the infection rate. Also, for this case, it was possible to solve the difference equation for the spreading analytically, exactly reproducing the ramped transition between the initial and final states of infection as well as the peaked profile of the infection rate. For permanently contagious molecules, the final fraction of infections was $\nu_f = 1$, as expected. Also, the width of the infection-rate was about 4 units of τ_s, the peak of the infection-rate profile was $0.25/\tau_s$, and the location of the peak was at $j \approx \log_2 N_0 + 0.471$, which is 17.08 for a population of 10^5 molecules.

In section 4, I dropped the assumption that the molecules were permanently contagious and derived the spreading equation for the fraction of infected molecules, ν_j, for several values of the contagious lifetime parameter, c. Immediately, the plots showed that this more general solution reduced to the previous one if c was greater than about 10. So if the molecules can stay red for at least 10 collision times, they will wipe out the entire population of blue molecules, as if these red molecules remained contagious forever.

However, for smaller values of c, and especially values less than 2 and approaching 1, the results differed quantitatively from the previous solution for $c > 10$. The trends were the same, showing the same ramped transition from the initial state of uninfected blue molecules to the final state when the epidemic was over. However, the difference was that the ramp of rising infection had a lower slope and a longer duration than when the red molecules had an unlimited lifetime. Also, the ramp started at a later time than when the red molecules were permanently contagious. Most important, the decrease in slope was greater than the increase in duration, causing their product (and the final level of infection) to be less than 1. At the end of the epidemic, all of the red molecules were gone. The population consisted entirely of green molecules that had lost their contagiousness and blue molecules that had escaped infection. These changes became more pronounced as c fell below 2 and approached 1, corresponding to the so-called ‘flattening of the curve‘ that happens with increased social distancing.

In the section on red, green, and blue molecules, we learned that the number of green molecules at a given time was equal to the total number of infected molecules at a time that is c steps earlier. This meant that the time history of green molecules is the same as that for all of the infected molecules, but delayed by c time steps. Both groups have the same three-phase evolution with an initially flat distribution, followed by a steep ramp of rapidly rising infections, which levels off at the final fraction of infected molecules. The difference between these two groups gives the fraction of red molecules. When the shift, c, is relatively small, this difference is just c times the derivative of the profile of infected molecules. Consequently, for small c, the fraction of red molecules has a profile that is just c times the profile for the growth rate of all the infected molecules (expressed mathematically, $\nu_{j-1} - \nu_{j-1-c} = c\Delta\nu_j$). For $c \geq 3$, this relation breaks down, and the rising phase of the profile for red molecules matches the ramped profile for all of the infected molecules. However, unlike the profile for the total number of infections (red plus green), the profile for red molecules reaches a peak and then returns to zero, leaving only green and uninfected blue molecules at the end.

Finally, in section 7, we examined the effect of changing the amount of social distancing during the epidemic. By suddenly removing all of the social distancing (*i.e.* increasing c to ∞, or, even to $c = 10$), all of the accomplishments disappeared and the fraction of infections rapidly shot up to the final value of $\nu_f = 1$. All of the the blue molecules became infected and the epidemic ended. For smaller changes either increases or decreases, the result was not quite so dramatic, but the trend was clear. To minimize the total number of infections, social distancing should be applied as soon as possible when entering the epidemic and it should be removed as late as possible when coming out of the epidemic.

The quantitative results, obtained from section 5 and appendices A-C, are:

1. The final fraction of infected molecules, ν_f was well fit by the exponential expression

$$\nu_f \;=\; 1 \;-\; e^{-1.89(c-1)}. \tag{8.1}$$

2. The 'width', w_e, defined by ν_f/h (where h is the maximum value of the infection rate), could be represented by the relation

$$\frac{1}{w_e} \;=\; 0.25[1 - e^{-0.806(c-1)}]. \tag{8.2}$$

The full-width at half maximum (FWHM) is somewhat less, $\sim 0.8 w_e$, depending on the detailed shape of the profile.

3. The peak height of the infection rate was given by

$$\Delta\nu_{max} \;=\; h \;=\frac{\nu_f}{w_e} \;=\; 0.25\left\{1 - e^{-0.806(c-1)}\right\}\left\{1 \;-\; e^{-1.89(c-1)}\right\}. \tag{8.3}$$

4. The location, j_{max}, of the peak height can be obtained from the empirical relation $\nu_j \approx \nu_f$, where ν_j is the dominant term in the solution of the linearized form of Eq.(4.4)

given by

$$\nu_j \approx \sigma_c \rho_c^j \, \nu_0. \tag{8.4}$$

Here, σ_c and ρ_c can be obtained from Table 1 of Appendix A or from the best-fit approximations given in Eqs.(8.7a) and (8.7b) below. Substituting $\nu_j = \nu_f$, and setting $j = j_{max} - 1$, we obtain

$$j_{max} - 1 \;=\; \frac{\log \nu_f}{\log \rho_c} \;-\; \frac{\log \sigma_c}{\log \rho_c} \;+\; \frac{\log N_0}{\log \rho_c}. \tag{8.5}$$

For c in the range (1,2), σ_c and ρ_c are given by

$$\sigma_c \;=\; \left(\frac{1}{c-1}\right)\left(1 + \frac{c}{s}\right), \tag{8.6a}$$

$$\rho_c \;=\; \left(\frac{1+s}{2}\right) \tag{8.6b}$$

where $s = \sqrt{1 + 4(c-1)}$. In addition, the best-fit approximations can be used over the full range of c:

$$\rho_c \;=\; 2 - e^{-0.920(c-1)}, \tag{8.7a}$$

$$\sigma_c \;=\; \left(\frac{2}{c-1}\right) e^{-0.372(c-1)} \;+\; \left\{1 - e^{-1.019(c-1)}\right\}. \tag{8.7b}$$

For $c > 3$, $\nu_f \approx 1$, $\sigma_c \approx 1$, and $\rho_c \approx 2$, so that Eq.(8.5) reduces to

$$j_{max} - 1 \;=\; \frac{\log N_0}{\log 2} \;=\; \log_2 N_0, \tag{8.8}$$

which is the result when the red molecules are permanently contagious.

5. A point-slope approach gave an alternate expression for j_{max}. In this approach, all of the linear plots of j_{max} versus $\log_{10} N_0$ were assumed to pass through the same point, $(\log_{10} N_0, j_{max}) = (1.586, 6.153)$, so that the expression for j_{max} became

$$j_{max} \;=\; 6.153 \;+\; \left(\frac{1}{\log_{10} \rho_c}\right)(\log_{10} N_0 - 1.586), \tag{8.9}$$

which reduces to

$$j_{max} - 0.884 \;=\; \frac{\log N_0}{\log 2} \;=\; \log_2 N_0, \tag{8.10}$$

when c is large and $\rho_c \approx 2$.

6. The starting position or threshold value, j_{th}, of the ramp of rapidly rising infection was defined by $\nu_{j_{th}} = 0.1\nu_f$, where $\nu_{j_{th}}$ is obtained from Eq.(8.4) with $j = j_{th}$, which is valid during the initial phase of the variation when $\nu_j << 1$. Consequently, j_{th} becomes

$$j_{th} \;=\; \frac{\log \nu_f}{\log \rho_c} \;-\; \frac{\log \sigma_c}{\log \rho_c} \;+\; \frac{\log N_0}{\log \rho_c} \;-\; \frac{1}{\log_{10} \rho_c} \;=\; (j_{max} - 1) - \frac{1}{\log_{10} \rho_c}. \tag{8.11}$$

Thus, $1/\log_{10}\rho_c$ gives the time interval between j_{th} and $j_{max}-1$, the place that the curve of accumulated infections has its maximum slope. (By comparison, j_{max} refers to the location of the maximum growth rate in the plot of $\Delta\nu_j$ versus j.)

7. The best-fit relation for the total fraction of infected molecules at the 'herd immunity' threshold was found to be

$$\nu_{herd} \;=\; 1 \;-\; e^{-0.860(c-1)}, \tag{8.12}$$

and the corresponding expression for the lag between the peak of the total infection rate, $\Delta\nu_j$ and the peak in the fraction of contagious (red) molecules, $\nu_j - \nu_{j-c}$, was found to be

$$j_{red} \;-\; j_{max} \;=\; 0.512(c-1) \;-\; 0.033(c-1)^2, \tag{8.13}$$

until the lag saturated at $c_m = 8.81$ and $j_{red} - j_{max} = 2$.

9. Discussion

Now, we will use these calculations to understand the propagation of a virus. Remember that they are for an idealized model of a gas of molecules and not for a realistic population of people in our society. Also, this model and its calculations are only a first step in the experiment. A second step might include more complicated interactions including multiple collisions, clumps of molecules, and exchanges of molecules between separate populations. On the other hand, this simple model is general and ought to apply to any virus and population for specified choices of the parameters, c (equivalently, τ_c), collision time τ_s, and N_0.

We found that when the infected molecules retained their contagiousness indefinitely and there was no 'social distancing', the number of infections increased exponentially until about half of the molecules were infected. After that, the number of infections increased more slowly as the remaining molecules of the gas became infected. The infection rate, which we represented by $\Delta\nu_j$, was essentially the derivative of the growth curve. It rose to its maximum value of $0.25N_0/\tau_s$ in a time on the order of $\log_2 N_0$, which is the number of powers of 2 in the number N_0 (equivalently, $3.322 \log_{10} N_0$).

This means that the time to infect half the population would depend on the size of the population, and therefore would be longer for a metropolis than for a small community. In particular, for a small town of 10^4 molecules, it would take 13.3 collisions to infect half the population, which would be 2-4 weeks if the average time between collisions were 1 or 2 days. On the other hand, for a large city of 10^6 molecules, it would take 19.9 collisions and therefore about 3-6 weeks.

On the other hand, we found that the shape of the infection-rate curve was independent of N_0, so that it had the same width and height for a small town and a large metropolis. In particular, its width (full width at half maximum) was about 4 collision times, which would be 4-8 days. Because $0.25N_0/\tau_s \times 4\tau_s = N_0$, this means that most of the molecules would be infected during this 4-step interval (of 4-8 days) around the peak of the distribution.

These results are a consequence of using Eq.(3.13a) to calculate ν_j, the number of infected molecules expressed as a fraction of the population, N_0. In this case, the final approach from $\nu = 0.5$ to $\nu = 1$ was slightly shorter than the initial rise from $\nu = \nu_0$ to $\nu = 0.5$. Equivalently, the infection-rate curve was asymmetric around its peak with a steeper fall than rise, which means that the epidemic ends faster than it starts.

Next, I discarded the assumption that the infected molecules remain contagious indefinitely, and solved the problem for a range of contagious lifetimes, τ_c, where τ_c is c times the collision time (or step time), τ_s. The result was a range of solutions that depend on the parameter, c. One way to think of this is to suppose that τ_s is a constant and τ_c varies, as if we were considering a variety of different diseases. Another approach is to suppose

that τ_c is constant, as it might be for a specific disease, and to consider what happens for a range of step times, τ_s that would occur for different amounts of 'social distancing'. So, for the current pandemic, we can think of the c-dependence as an indication of the influence of 'social distancing' on the spread of the disease with $c = 1$ corresponding to perfect distancing and $c = \infty$ corresponding to the normal society with no social distancing.

The first result of these new calculations was to find that the plots of the number of infected molecules (expressed as a fraction of the total population, N_0) were similar to the original plot with no social distancing, showing a relatively long base level followed by a steep ramp to the final level where the infection stops. However, for increased amounts of social distancing, the curves took longer to reach the steep ramp, the slope of the ramp was lower, and the final level of infected molecules was smaller than in the absence of social distancing. Likewise the growth-rate profiles had smaller heights, larger widths, and were shifted to later times as the amount of social distancing increased. This corresponds to the popular term, 'flattening the curve'.

But how much flattening and delaying occurred? Let us assume that the contagious time is $\tau_c = 14$ days based on the time that an infected person is required to be quarantined for the current corona virus. As shown in Figure 7, for $c = 10$ (corresponding to $\tau_s = \tau_c/c \approx 14/10 = 1.4$ days between collisions), the curve peaks at $j_{max} = 17$, corresponding to $1.4 \times 17 = 23.8$ days after the first contagious molecule entered the population of $\sim 10^5$ molecules. For $c = 1.25$ (corresponding to $14/1.25 = 11.2$ days between collisions), the curve peaks at $j_{max} = 48$, corresponding to $11.2 \times 48 = 537.6$ days, or approximately 18 months. Referring to Figure 6, we found that the fraction of infected molecules dropped from 1.0 to 0.38 (100,000 dropped to 38,000). So this social distancing saved 62,000 blue molecules, but required more than 1.5 years to do it. And this refers to the peak height, not the final infection rate, which reaches 0 around $j_{max} = 70$, corresponding to $11.2 \times 70 = 784$ days, which is about 26 months. So the cost of those 62,000 blue molecules was more than 2 years of keeping apart and avoiding collisions for intervals of about 11 days.

The times are shorter for smaller populations and longer for larger populations, as we found in Eq. (5.10) and Figure 12. For a small town of population $N_0 = 10^4$, the peaks occur after times of $1.4 \times 13.6 = 19$ days for $c = 10$ and $11.2 \times 35.5 = 397.6$ days (about 13 months) for $c = 1.25$. For a metropolis with $N_0 = 10^6$, the corresponding times are $1.4 \times 20.2 = 28.3$ days and $11.2 \times 60.2 = 674$ days (about 22 months). And these are only the times to reach the peak infection rates. The time required to bring those rates back to 0 is even larger. The number of blue molecules saved by social distancing is proportional to the populations, so for the smaller population of 10^4, the number saved is 6,200, and for the metropolis of 10^6, the number saved is 620,000.

As we have seen in Figure 6, the growth of the numbers of infected molecules seems to go through three phases. The evolution begins with a long interval of low values, eventually

changing to an upward ramp that ends in a plateau of final values. The width of the infection-rate profile is a rough measure of the duration of this upward ramp, and therefore an indication of how long the infection-rate remains at high values. As shown in Figure 7 and plotted in Figure 10, the width, $w_e = \Delta j$, decreases as c increases. Therefore, the width increases as the social distancing increases. In fact, from Eq. (5.8a), one can show that the width, w_e, is given approximately by $w_e = 5/(c-1)$ for a large amount of social distancing with $c \lesssim 1.5$.

Thus, for the example with $c = 1.25$ ($\tau_s \approx 11.2$ days), the width, w_e, is $20 \times 11.2 = 224$ days, or about 7 months. Whereas, for our previous example with $c = 10$ ($\tau_s = 1.4$ days), the width was approximately $4 \times 1.4 = 5.6$ days. So when social distancing reduces the fraction of infected molecules from 1 to 0.38, it increases the duration of the active phase of the disease from about 1 week to 7 months. By social distancing, we trade a short duration of infection at a high rate for a long duration of infection at a low rate. This is the so-called 'flattening of the curve'. However, as Eq.(5.9a) shows, the benefit is to substantially reduce the final number of infected molecules.

It is also interesting to understand this process in terms of the numbers of contagious molecules (expressed as a fraction of the total population, N_0). For example, if a particular molecule managed to escape the virus and remain blue for a while, what would its chances of encountering a contagious molecule be at that time? Recall that the number of contagious molecules at a given step time, j, is equal to the number of infected molecules at that time minus the number of infected molecules c steps earlier at the time $j - c$. This difference depends on the size of the shift, c, relative to the width, w, of the infection-rate profile. For social distancing with $c \leq 2$, the peak number of contagious molecules is approximately c times the maximum slope of the growth curve (which is the peak value of the infection-rate curve). So the resulting number of contagious molecules in the top panel of Figure 17 is just c times the corresponding peak heights in Figure 7. Because $c > 1$ for an uncontained virus, this means that the peak heights for the numbers of contagious molecules are always greater than the corresponding peak heights for the infection-rate profiles. On the other hand, for a nominal amount of social distancing, the widths and locations of the peaks are the same for the contagious (red) molecules as they are for the infected (red plus green) molecules. Based on the numbers that we found in the examples above, this means that the surviving blue molecules must wait a few years before re-entering the pool if they wish to avoid collisions with contagious red molecules.

It is important to remember that this molecular model and its numerical calculations began as an attempt to determine the mathematical properties of the growth curve and growth-rate curve. The model was not intended to be a realistic description of the spread of the Covid-19 disease. Even the choice of 14 days for τ_c was done arbitrarily to permit the calculation of some specific delay times that might occur in the evolution of the disease. The resulting delays of months and years might be different if another value of τ_c were used

in the calculation. Nevertheless, the reader should be pleased to gain some insight into how the process works, and not be too alarmed or complacent about the resulting numbers. Remember, these numbers refer to molecules, not people.

Like many people, I was interested in how a virus might spread. Rather than looking for the answer in the published literature, I created the molecular model and did the calculations as if I were solving a puzzle in math or physics. Only after finishing these calculations, did I perform a Google search and find the three-component model of Kermack and McKendrick (1927) and a summary of that model by Weisstein (2004). Of course, this was 'the tip of the iceberg', which led to more recent references, including those of Anderson and May (1979), Jones and Sleeman (1983), Smith and Moore (1996), and others. The three components of the 'SIR-model' referred to individuals who were susceptible (S), infected and contagious (I), and recovered and not contagious (R) - the same components that I have called blue (B), red (R), and green (G), respectively. Therefore, we should probably call our molecular model the RGB model, after the RGB color coding used in electronic imaging and photography. Kermack and McKendrick described these populations with three differential equations, of which only two are independent because the total number of molecules is conserved. In our RGB terminology, $R + B + G = N_0$, and the SIR equations are:

$$\frac{d(R+G)}{dt} = +\beta BR, \tag{9.1a}$$

$$\frac{dR}{dt} = \beta BR - \gamma R, \tag{9.1b}$$

$$\frac{dG}{dt} = \gamma R. \tag{9.1c}$$

Here, $1/\beta$ corresponds to the collision time, τ_s, and $1/\gamma$ is the $(1/e)$-lifetime of the red molecules in the absence of a source, as one can see by setting $B = 0$ in Eq.(9.1b). (Equivalently, $1/\gamma$ is the average lifetime of the red molecules.) In the RGB-model, the infected molecules remain contagious for a finite time $\tau_c = c\tau_s$ before suddenly losing their ability to infect. Thus, c is equivalent to the average lifetime, $1/\gamma$, in units of the collision time, $1/\beta$.

As Weisstein pointed out, the key value determining the propagation or damping of these equations is the 'epidemiological parameter', $R_0 = (\beta/\gamma)B$, where B stands for the fraction of uninfected blue molecules, and β/γ is analogous to the quantity that we call c in the RGB-model. Thus, in Eq.(9.1b), dR/dt changes sign from positive to negative when $(\beta/\gamma)B$ falls below 1, and the number of red molecules starts to decrease. Setting $B = 1$, we obtain $R_0 = \beta/\gamma$, which is the fraction of molecules that a single red molecule at the start of the epidemic will infect before it turns green. However, in the RGB-model, that red molecule would be responsible for $2^c - 1$ infections because each infected molecule goes on to infect other blue molecules before the original red molecule suddenly turns green. With this distinction, c is analogous to R_0, and the difference equations of the RGB-model are analogous to the differential equations of the SIR-model, as will be described in detail in Appendix D.

A. Solution For $\nu_j << 1$

We have seen that when $c \geq 10$, the fraction of infected molecules, ν_j, can be obtained from Eq. (3.7), which has an exact solution given by $\nu_j = 1 - 2^{-2^{\{j-(jmax-1)\}}}$, where j_{max} is where the growth rate $\Delta\nu_j = \nu_j - \nu_{j-1} = 1/4$ and $\nu_{j-1} = 1/2$. This is the solution when the infected molecules remain contagious for the duration of the epidemic, and all of the blue molecules eventually become red. Also, recall that when the non-linear term in Eq. (3.7) is neglected, the linearized equation becomes $\nu_j = 2\nu_{j-1}$ and its solution is $\nu_j = 2^j\nu_0$. This raises the question of what the initial-phase of the solution is for other values of c.

To find out, we linearize Eq. (4.3) by replacing the factor $1 - \nu_{j-1}$ by 1 and writing

$$\nu_j \;=\; 2\nu_{j-1} \;-\; \nu_{j-1-c}. \tag{A1}$$

This linear equation has an exact solution composed of terms of the form, r^j, where r refers to the roots of the equation, $r^{c+1} - 2r^c + 1 = 0$, obtained by substituting r^j into Eq. (A1). When $c = 2$, this equation is $r^3 - 2r^2 + 1 = 0$, whose three roots are $r = 1, (1 \pm \sqrt{5})/2$. Consequently, when $c = 2$, the exact solution to Eq. (A1) is

$$\frac{\nu_j}{\nu_0} \;=\; -1 \;+\left(1 + \frac{2}{\sqrt{5}}\right)\left\{\frac{1+\sqrt{5}}{2}\right\}^j \;+\; \left(1 - \frac{2}{\sqrt{5}}\right)\left\{\frac{1-\sqrt{5}}{2}\right\}^j, \tag{A2}$$

where the coefficients, $-1, 1 \pm 2/\sqrt{5}$ were determined by setting $\nu_j/\nu_0 \approx 1, 2, 4$ for $j = 0, 1, 2$ and solving the three equations simultaneously. The approximate numerical form of Eq. (A2) is

$$\frac{\nu_j}{\nu_0} \;=\; -1 \;+ 1.894(1.618)^j \;-\; 0.106(-1)^j(0.618)^j, \tag{A3}$$

which is quickly dominated by $1.894(1.618)^j$ when j becomes larger than about 6. This is to be compared with $1.355(1.839)^j$ when $c = 3$ and 2^j when c is greater than 10. Also, as we shall see in the next few paragraphs, the dominant terms are $3.732(1.366)^j$ for $c = 1.5$ and $7.536(1.207)^j$ for $c = 1.25$. These results are given in Table 1 for an expression of the form

$$\frac{\nu_j}{\nu_0} \;=\; \sigma_c\ {\rho_c}^j, \tag{A4}$$

where σ_c is the coefficient of the dominant r^j-term, and ρ_c is the corresponding value of r. As we would expect, this dominant term approaches 2^j as c becomes large. And as we will show next, it approaches $\{2/(c-1)\}c^j$ as $c \to 1$.

We are especially interested in the solution when c lies between 1 and 2. Recall that we used a linear interpolation for ν_{j-1-c} in section 4. For c in the range (1,2), this linear interpolation is $\nu_{j-1-c} = (2-c)\nu_{j-2} + (c-1)\nu_{j-3}$. Consequently, Eq. (A1) becomes

$$\nu_j \;-\; 2\nu_{j-1} \;+\; (2-c)\nu_{j-2} \;+\; (c-1)\nu_{j-3} \;=\; 0, \tag{A5}$$

Table 1: σ_c and ρ_c

c	σ_c	ρ_c
1.25	7.536	1.207
1.50	3.732	1.366
1.75	2.500	1.500
2.00	1.894	1.618
2.50	1.694	1.740
3.00	1.355	1.839
4.00	1.177	1.928
5.00	1.095	1.966
∞	1.000	2.000

whose associated equation for r is

$$r^3 \; - \; 2r^2 \; + \; (2-c)r \; + \; (c-1) \; = \; 0. \tag{A6}$$

The three roots of this equation are $r = 1, \{1 \pm \sqrt{1+4(c-1)}\}/2$. As before, we set $\nu_j/\nu_0 =$ 1, 2, and 4 for $j = 0$, 1, and 2, and solve for the coefficients of each r^j term.. Then, with the substitution $s = \sqrt{1+4(c-1)}$, we can write the general solution for ν_j/ν_0 as

$$\frac{\nu_j}{\nu_0} \; = \; \left(\frac{c-3}{c-1}\right) \; + \; \left(\frac{1}{c-1}\right)\left(1+\frac{c}{s}\right)\left\{\frac{1+s}{2}\right\}^j \; + \; \left(\frac{1}{c-1}\right)\left(1-\frac{c}{s}\right)\left\{\frac{1-s}{2}\right\}^j, \tag{A7}$$

valid for $1 < c \leq 2$. It is reassuring to see that when $c = 2$, $s = \sqrt{5}$, and Eq. (A7) reduces to Eq. (A2). At the other end of the range where $c - 1 << 1$, Eq. (A7) approaches the relation

$$\frac{\nu_j}{\nu_0} \approx \; -\left(\frac{2}{c-1}\right) \; + \; \left(\frac{2}{c-1}\right)c^j \; + \; (-1)^j(c-1)^j, \tag{A8}$$

which allows us to see that ν_j/ν_0 is dominated by $\{2/(c-1)\}c^j$ when c is close to 1 and $c^j >> 1$. More generally, we can approximate Eq. (A7) by

$$\frac{\nu_j}{\nu_0} \approx \; \left(\frac{1}{c-1}\right)\left(1+\frac{c}{s}\right)\left\{\frac{1+s}{2}\right\}^j, \tag{A9}$$

where $s = \sqrt{1+4(c-1)}$ and $1 < c \leq 2$.

We can use Eq.(A4) to compare this analytical solution of the linearized Eq.(A1) with the numerical solution of the non-linear Eq.(4.3). The results are plotted in Figure 25 for $c = 10, 2, 1.5, 1.25$ and $N_0 = 1 \times 10^5$. (When $c = 10$, the linearized equation is $\nu_j = 2\nu_{j-1}$, whose solution is $\nu_j = 2^j \nu_0$.) Recall that the linearized equations involve the approximation $1 - \nu_{j-1} << 1$, which means that we expect the approximate solutions to be valid only for small ν_j. In Figure 25, we see that the solutions for $c = 10$ and $c = 2$ are fairly accurate for values of ν_j less than about 0.2. However, for progressively smaller values of c, the dashed and solid curves begin to separate at smaller values of ν_j, below 0.1 for $c = 1.5$ and then below 0.05 for $c = 1.25$. However, the values of ν_f also decrease as c becomes closer to 1, so the discrepancies will be smaller when normalized to the values of ν_f.

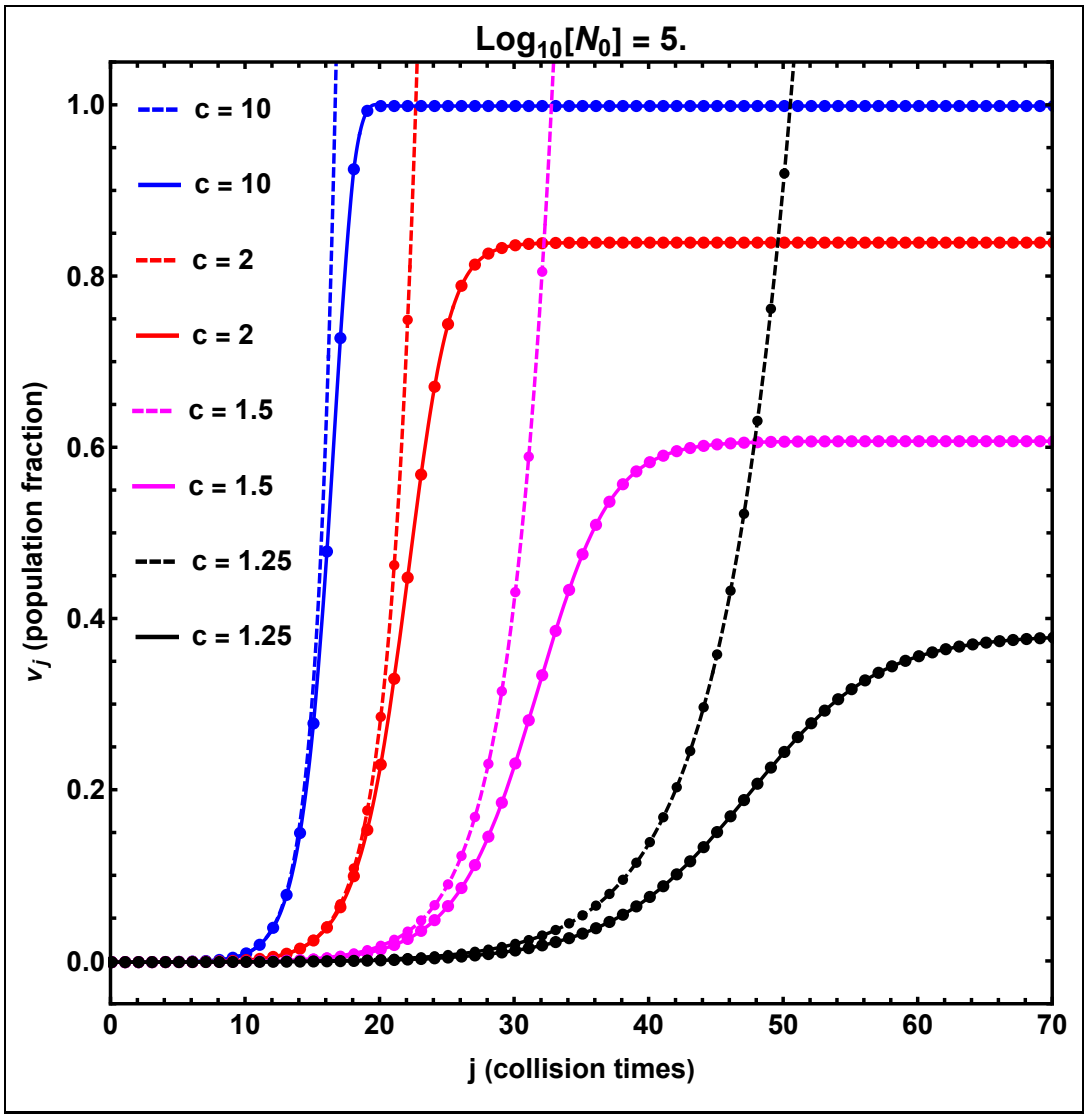

Fig. 25.— Comparing the solutions of the linearized equation (dashed) with the solutions of the non-linear equation (solid) for the growth of the virus when the values of c are 10, 2, 1.5, and 1.25, and when $\log_{10} N_0 = 5$.

B. Estimating the Starting Point of the Region of Rapid Increase

Although the solution of the linearized version of Eq. (4.3) diverges from the solution of the complete non-linear equation, it remains accurate near the initial rising phase of the evolution where $\nu_j << 1$. Consequently, we can use the solution given by Eq.(A4) to estimate where the rapid increase of ν_j begins. We begin by supposing that the starting point can be defined by a threshold value, arbitrarily taken to be $\nu_j = 0.1\nu_f$. The corresponding value of j, say j_{th}, is determined by the equation

$$\sigma_c \rho_c^{j_{th}} \;=\; 0.1(\frac{\nu_f}{\nu_0}) \;=\; 0.1\;\nu_f N_0. \tag{B1}$$

Next, we set $j = j_{max} - 1$ in Eq.(5.11) and divide that equation by Eq.(B1). This division eliminates σ_c, ν_f, and N_0, and gives

$$\rho_c^{j_{max}-1-j_{th}} \;=\; 10, \tag{B2}$$

from which we deduce that

$$(j_{max} - 1) - j_{th} \;=\; \frac{1}{\log_{10}\rho_c}. \tag{B3}$$

Thus, we have an expression for the difference, $(j_{max} - 1) - j_{th}$, as a function of ρ_c, which is given by Table 1 of Appendix A and the solution to the equation $r^{c+1} - 2r^c + 1 = 0$ when c is an integer greater than 1. (Recall that $j_{max} - 1$ and j_{max} refer to different profiles. $j_{max} - 1$ is the location of the maximum slope of ν_j, which concerns us here, whereas j_{max} is the location of the peak of $\Delta\nu_j$.)

Note that our arbitrary threshold of $\nu_{th} = 0.1\nu_f$ led to the shift $(j_{max} - 1) - j_{th} = -\log_{10} 0.1/\log_{10}\rho_c = 1/\log_{10}\rho_c$ because the logarithms were taken to base 10. But we could have used any base, including ρ_c, in which case, the shift would have been $-\log_{\rho_c} 0.1/\log_{\rho_c}\rho_c = \log_{\rho_c} 10$. For permanently contagious red molecules, $\rho_c = 2$ and the shift becomes $\log_2 10 = 3.322$. For convenience, we could have chosen a threshold using a power of 2, like $\nu_{th} = 2^{-3}\nu_f$, for example, so that the shift is just $\log_2 2^3 = 3$ collision times. Thus, for permanently contagious red molecules, the threshold j_{th} would be just 3 steps back from the point of greatest slope, which itself is given approximately by $j_{max} - 1 \sim \log_2 N_0$.

Also, note that because the shift is $(j_{max}-1)-j_{th} = 1/\log_{10}\rho_c$, the shift will double from 3 to 6 collision times when ρ_c is changed from 2 to $\sqrt{2} = 1.414$ (corresponding to $c = 1.581$ using Eq. (C1a)). And it will double again to 12 collision times on going to $\rho_c = 2^{1/4} = 1.189$ ($c = 1.228$), as one can verify in Figure 6 (but remembering that the threshold value is not just 0.1 (or 1/8 for the simpler case), but is $0.1\nu_f$ (or $(1/8)\nu_f$). Likewise, for the nearly symmetric profiles, the end of the ramp occurs about the same number of collision times after the point of greatest slope.

C. Approximating the Linearized-Solution Parameters, σ_c and ρ_c

Because the parameters, σ_c and ρ_c are important for understanding the initial behavior of ν_j, I thought it would be useful to obtain approximate analytical expressions for those parameters, expressed as a function of c. Consequently, in Figures 26 and 27, I have plotted the values of ρ_c and σ_c, respectively, obtained from Table 1 of Appendix A, supplemented by additional values chosen to fill in the range of c. In addition, I have superimposed dashed curves corresponding to the best-fit exponential solutions given in Eqs. (C1a) and (C1b).

$$\rho_c \;=\; 2 - e^{-0.920(c-1)}, \tag{C1a}$$

$$\sigma_c \;=\; \left\{\frac{2}{c-1}\right\} e^{-0.372(c-1)} \;+\; \left\{1 - e^{-1.019(c-1)}\right\}. \tag{C1b}$$

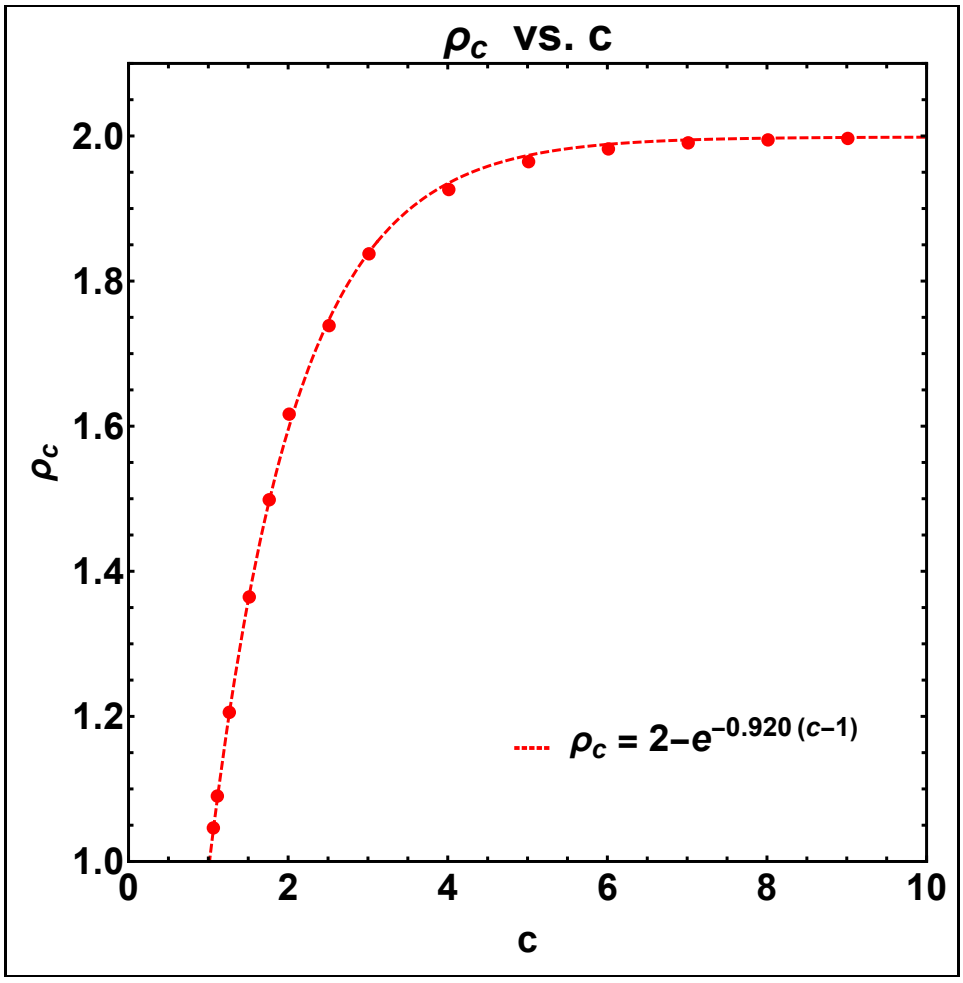

Fig. 26.— Calculated values of ρ_c (red points), superimposed by the best-fit solution given by Eq.(C1a) (dashed red curve). ρ_c is defined by $\nu_j = \sigma_c\ \rho_c^j\ \nu_0$, the dominant term in the solution to the linearized equation of spreading.

The general form of Eq.(C1a) was chosen so that it would give 1 when $c = 1$ and

give 2 as $c \to \infty$. That was much easier than choosing Eq.(C1b) so that it would reduce to $2/(c-1)$ as $c \to 1$ and approach 1 as $c \to \infty$. An additional exponential factor was necessary to weaken the $2/(c-1)$-term so that σ_c would approach 1 more rapidly than $2/(c-1)$ as $c \to \infty$.

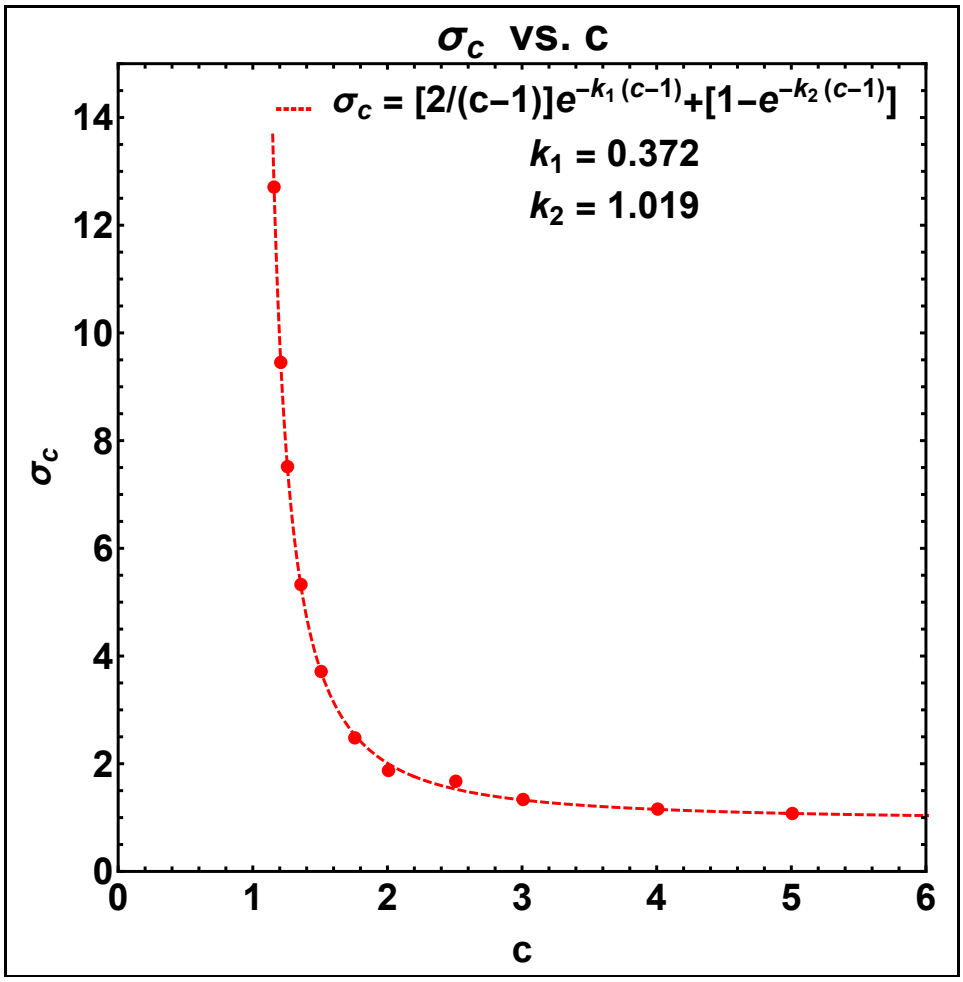

Fig. 27.— Calculated values of σ_c (red points), superimposed by the best-fit solution given by Eq.(C1b) (dashed red curve). σ_c is defined by $\nu_j = \sigma_c\ \rho_c^j\ \nu_0$, the dominant term in the solution to the linearized equation of spreading.

Finally, note that σ_c contributes to the initial behavior of ν_j as a shift in the value of j. We can see this by writing

$$\nu_j \ = \ \sigma_c\ \rho_c^j \ = \ \rho_c^{j+j_{sh}}, \tag{C2}$$

where $j_{sh} = \log \sigma_c / \log \rho_c$ whose logarithms can be taken to any base. In this case, Eq. (5.12) of the main text can be rewritten as

$$j_{max} - 1 \ = \ \frac{\log \nu_f}{\log \rho_c} \ + \ \frac{\log N_0}{\log \rho_c} \ - \ j_{sh}, \tag{C3}$$

Thus, if c were greater than about 2, ν_f and σ_c would be approximately 1, ρ_c would be approximately 2, and j_{sh} would be negligible. Eq.(C3) would then become

$$j_{max} - 1 \;=\; \frac{\log N_0}{\log 2} \;=\; \log_2 N_0, \tag{C4}$$

which is the result for molecules that remain contagious indefinitely.

Figure 28 shows values of j_{sh} plotted as a function of c. For comparison, the dashed blue curve indicates j_{sh}, calculated from Eq.(A9) and $j_{sh} = \log \sigma_c / \log \rho_c$. Eq.(A9) is the approximate solution of Eq.(A5), which only applies when c lies in the interval (1,2). Nevertheless, the extension of this curve fits the data points for $c > 2$ reasonably well. A horizontal line of the form $j_{sh} = 0$ would also be a good fit for $c > 3$. Consequently, we could also fit the data points using Eq.(A9) inside (1,2) and $j_{sh} = 0$ for $c > 3$ outside.

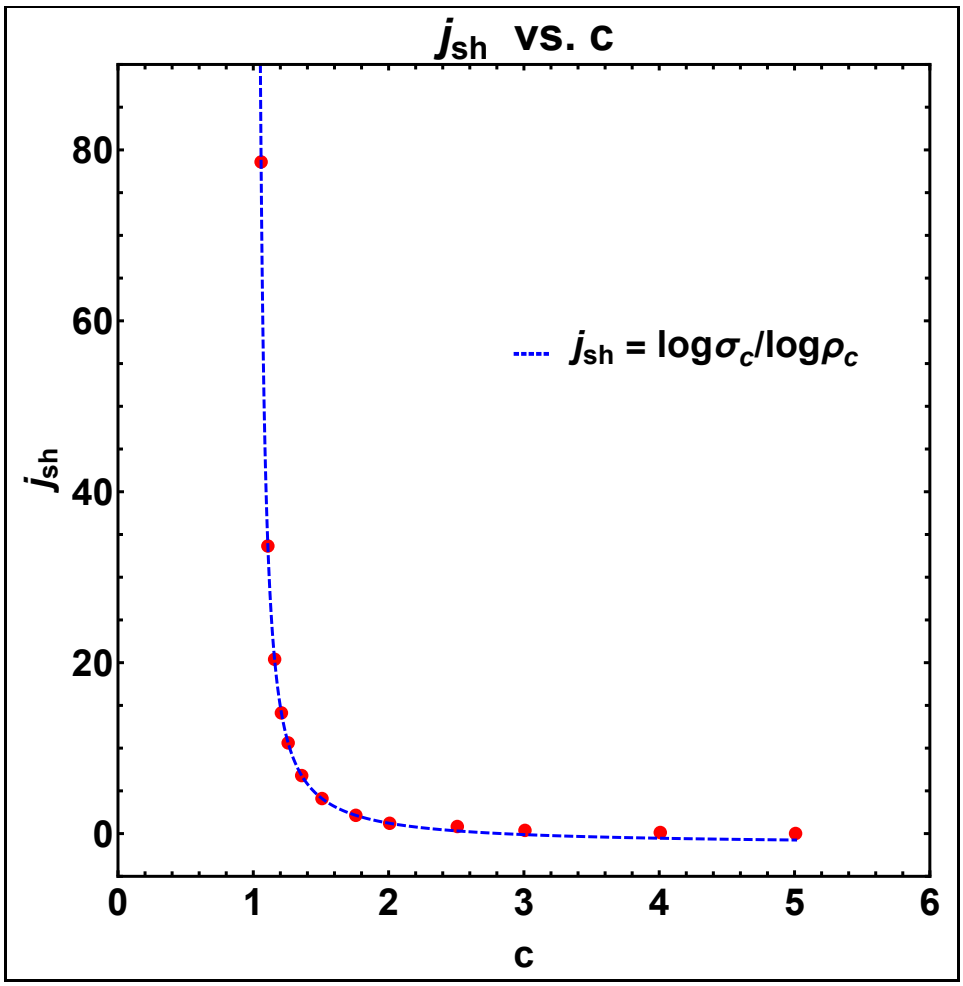

Fig. 28.— Calculated values of j_{sh} (red points) superimposed by a plot of Eq.(A9), the approximate solution of Eq.(A5) (dashed blue line). Although Eq.(A5) only applies for c in the interval (1,2), its extension fits the nearly constant values of $j_{sh} = 0$ for $c > 3$ very well.

D. Comparison with the Solutions of the SIR Equations

As I mentioned in the discussion section of this article, I constructed the RGB-model and solved its equations without any reference to the published literature. I just wanted to see if I could figure it out on my own, as if one of my scientific friends had presented it to me as a mathematics puzzle. It was only afterward, when I started looking through the published literature that I found the well-known SIR-model of Kermack and McKendrick (1927).

Both models describe the spread of a virus through a large population of initially uninfected individuals, and track the population of the same three types of individuals, which I have called red (infected and contagious), blue (uninfected), and green (infected, but no longer contagious). However, the models take different approaches. Whereas the RGB-model is based on individual collisions and discrete equations, the SIR-model is continuous and is controlled by differential equations. The contrast is like comparing the discrete spreading by a random walk with the continuous spreading by an effective diffusion, reminiscent of the two ways of simulating the transport of magnetic flux on the Sun's surface (Leighton, 1964; Wang and Sheeley, 1994).

Although both models suppose that the infected individuals do not remain contagious indefinitely, the RGB-model turns off the contagiousness abruptly after a specific time, τ_c, whereas the SIR-model allows the contagiousness to decay exponentially at a rate γ. The models have two other relevant parameters, τ_s, the step rate or average time between collisions of the molecules in the RGB-model, and β, the rate at which the contagious red molecules and uninfected blue molecules come together. In each case, the important quantity is the ratio of these respective parameters – τ_c/τ_s for the RGB-model and β/γ for the SIR-model. I have used c for τ_c/τ_s, and in this section, I will also use c for the ratio, β/γ.

We begin with the SIR equations Eqs.(9.1a), (9.1b), and (9.1c) in the discussion section of this paper. They are subject to the initial conditions: $R(0) = \epsilon = 1/N_0$, $G(0) = 0$, and $B(0) = 1 - \epsilon$. This corresponds to our starting condition of one red molecule and $N_0 - 1$ blue molecules in the RGB-model. In each case, $R + G + B = 1$ in our normalized units.

Our first step will be to relate the quantities, R, R+G (which corresponds to ν_j in the RGB-model), and B, to the variable, G. We do this by dividing Eq.(9.1a) by Eq.(9.1c) to obtain

$$\frac{d(R+G)}{dG} \;=\; \frac{\beta}{\gamma}\,[1-(R+G)]\,. \tag{D1}$$

Inserting $c = \beta/\gamma$ and solving this differential equation, we obtain

$$R + G \;=\; 1 \;-\; (1-\epsilon)e^{-cG}, \tag{D2}$$

where we have chosen the constant of integration to satisfy the initial conditions of the previous paragraph. And, of course, once $R + G$ is known, B is also determined because their sum is 1:

$$B \;=\; (1-\epsilon)e^{-cG}. \tag{D3}$$

It takes a little more algebra to obtain R. We begin by rewriting Eq.(D1) as

$$\frac{dR}{dG} \;+\; cR \;=\; (c-1) \;-\; cG. \tag{D4}$$

The solution to this differential equation is

$$R \;=\; 1 \;-\; G \;-\; (1-\epsilon)e^{-cG} \tag{D5}$$

where we have again determined the constant of integration so that the resulting equation satisfies the initial conditions of the first paragraph. Now, we have R, $R + G$, and B expressed in terms of G. Unfortunately, that is the end of our analytical solutions. Next, we use numerical integration to evaluate G as a function of time, t.

We begin by combining Eq.(9.1c) and Eq.(D5) as

$$\frac{dG}{dt} \;=\; \gamma R \;=\; \gamma\left[1 \;-\; G \;-\; (1-\epsilon)e^{-cG}\right], \tag{D6}$$

from which we can obtain an implicit expression for $G(t)$:

$$t \;=\; c\int_0^{G(t)} \frac{dG}{1-G-(1-\epsilon)e^{-cG}}, \tag{D7}$$

where again we have adjusted parameters, this time by setting $\gamma t = (\gamma/\beta)(\beta t) = (1/c)(t/\tau_s)$ and expressed t in units of τ_s.

As a practical point, we note that the presence of ϵ prevents the denominator of the integrand from vanishing at the lower limit of $G = 0$. It also provides the dependence on the population, $N_0 = 1/\epsilon$. As a second point, we note that the upper limit, $G(t)$, is bounded by a maximum value, which is 1 when $c = \infty$, but less than 1 for finite values of c. Because $R = 0$ at the end of the epidemic, it follows that the maximum value of G is also the maximum value of $R+G$, which we have called ν_f in the previous sections. In this case, the maximum value of G is determined by setting the denominator of the integrand equal to zero.

This provides a method for calculating the value of ν_f in this SIR-model. We just set $1 - G - (1-\epsilon)e^{-cG} = 0$, neglect ϵ compared to 1, and solve for c as a function of G. The result is

$$c \;=\; -\frac{ln(1-G_f)}{G_f} \;=\; -\frac{ln(1-\nu_f)}{\nu_f}, \tag{D8}$$

where G_f is the final value of G and $G + R$, which corresponds to ν_j in the RGB-model. Figure 29 shows this value of $\nu_f(c)$ plotted as the black, dashed curve with the corresponding

values obtained from our RGB-model, already shown in Figure 8. The agreement is fairly good overall, and it is nearly perfect in the region where $c < 1.4$. As we shall see next, the agreement for small values of c is a general property of the two models. Equivalently, the disagreement increases as the exponential decay time for R increases in the SIR-model.

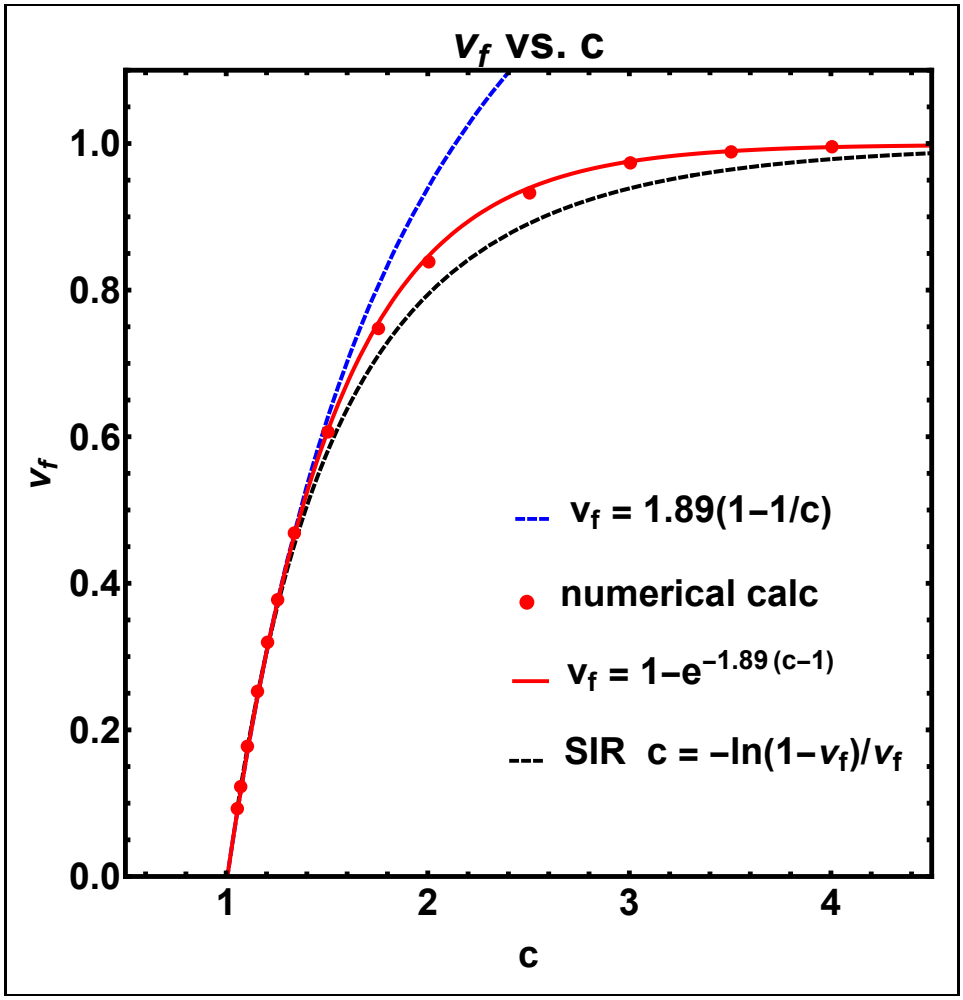

Fig. 29.— The final value, ν_f, of the growth profile plotted versus the contagious lifetime, c, comparing the results from the RGB-model (solid red and dashed blue curves) and the SIR-model (black dashed curve) (cf. Figure 8).

Next, we use Eq.(D7) to create a table of numerical values of $G(t)$, and then use Eqs.(D2) and (D5) to create corresponding tables of $R(t) + G(t)$ and $R(t)$. The results are plotted in Figure 30 for values of $c = 1.5, 2, 3$ and 10. Referring to the corresponding RGB-plots in Figure 15, we see again that the two models are similar for small values of c. In each model, for $c = 1.5$ and $c = 2$, the variation of G is essentially the same as the variation of $R + G$, but shifted c units later in time, and the variation of R is approximately c times the corresponding growth rate. This starts to change when $c = 3$, and by $c = 10$, $R + G$

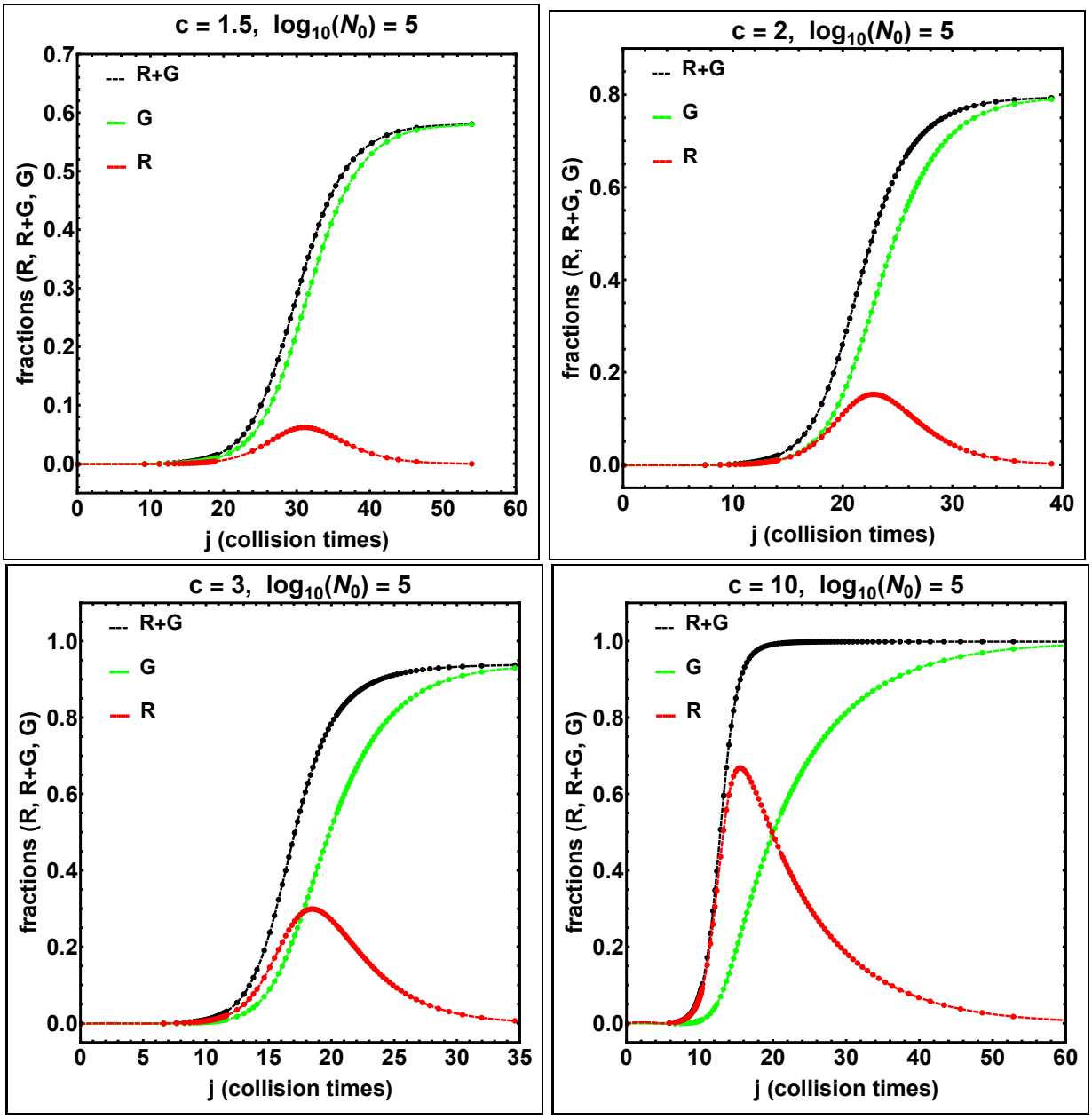

Fig. 30.— The temporal evolution of R (red curve), G (green curve), and $R+G$ (black curve) populations according to the SIR-model, for comparison with the corresponding evolutions in the RGB-model (cf. Figure 15). Note the scale changes between some of the images.

increases so fast that it reaches about 0.67 before G has begun its increase. By this time, R no longer clings to the $R+G$ curve, and begins its descent like an exponential of the form e^{-t/τ_c} while G shows the corresponding exponential increase, $1-e^{-t/\tau_c}$. For the RGB-model in Figure 15, when $c = 10$, G continues to be a delayed version of $R+G$. Likewise, R rises with $R+G$ until G starts to increase, and then falls to 0 as G increases to 1. Thus, when $c = 10$ or more, the ultimate decline of the number of red molecules reflects the steep

rise of the number of green molecules in the RGB-model, but it reflects the more gradual, exponential rise of the number of green molecules in the SIR-model.

The exponential behavior of the SIR model is shown even more clearly for the plots with $c = 20$ and $c = 50$ in Figure 31. In the SIR model, these plots show a sudden infection of the entire population followed by an exponential decrease of the red population as it gradually turns green on the time scale, τ_c.

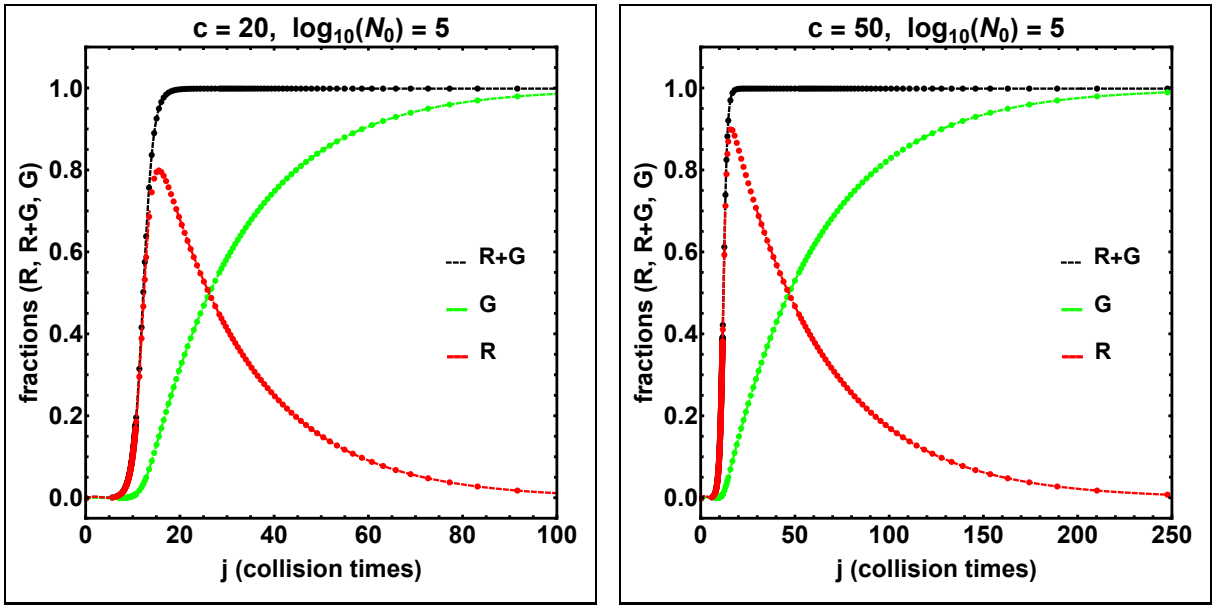

Fig. 31.— The extension of the SIR plots in Figure 30 to values of $c = 20$ and $c = 50$, showing the steepening of $R + G$ and the exponential appearance of G and R.

Another approach is to display the SIR plots of $R + G$ and $d(R + G)/dt$ in the same way that we displayed ν_j and $\Delta\nu_j$ in Figures 6 and 7. Figures 32 and 33 show those SIR plots. At first glance, the plots look remarkably similar. However, a closer look shows some differences. In Figure 32, the asymptotic values of $R + G$ are slightly less than the corresponding values of ν_f, as we have already seen in the plots of ν_f versus c in Figure 29. Also, in Figure 32, the 'knees' of the plots of $R+G$ versus time are more rounded than those for the RGB-model in Figure 6, especially for the larger values of c. This causes the SIR plots of the growth rate in Figure 33 to return to 0 more gradually than the corresponding RGB-plots in Figure 7. Likewise the peak heights for the SIR growth rates are slightly less than those for the RGB-model, except when $c < 2$. Again, this shows the general tendency for the models to agree when c is small and there is a lot of social distancing. As a final part of our comparison between the plots in Figures 32 and 33 and in Figure 6 and 7, we note that the SIR plots of the growth rate reach their maximum values a few steps before the corresponding RGB growth rates reach their peaks.

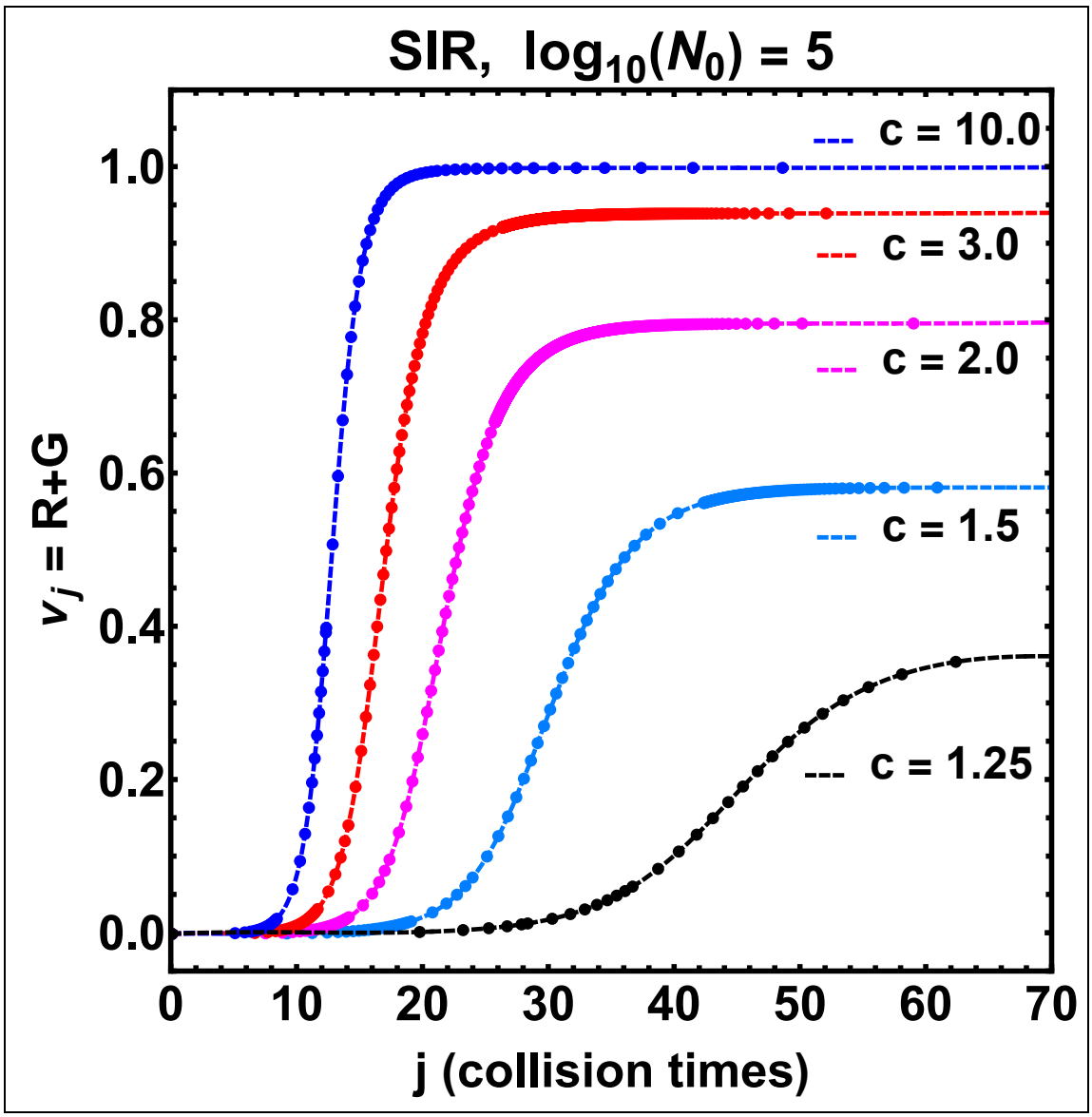

Fig. 32.— The fraction of infected population, $R+G$, calculated from the SIR-model using Eqs.(D2) and (D7), for values of $c = 1.25, 1.5, 2.0, 3.0$ and 10. These curves are analogous to the plots of ν_j in Figure 6 and show the delayed onset and the reduced final fraction of the infected population as c approaches 1.

In our discussion of 'herd immunity' in section 6.2, we defined the threshold to be the value of ν_j where the number of red molecules, $\nu_j - \nu_{j-c}$, has its maximum value and is starting to decline. The resulting expressions are much simpler in the SIR-model than in the RGB-model. We begin with Eq.(9.1c) which tells us that $dG/dt = \gamma R$, and therefore that $d^2G/dt^2 = \gamma dR/dt$. Thus, the herd-immunity condition, $dR/dt = 0$, implies that the slope, dG/dt, is a maximum when R is a maximum. (By way of contrast, in the RGB-model, R obtained its maximum value slightly before ν_{j-c} obtained its maximum slope.) As Eq.(9.1b) indicates, $dR/dt = \beta RB - \gamma R$, so that $dR/dt = 0$ also implies that $B = 1/c$ and therefore

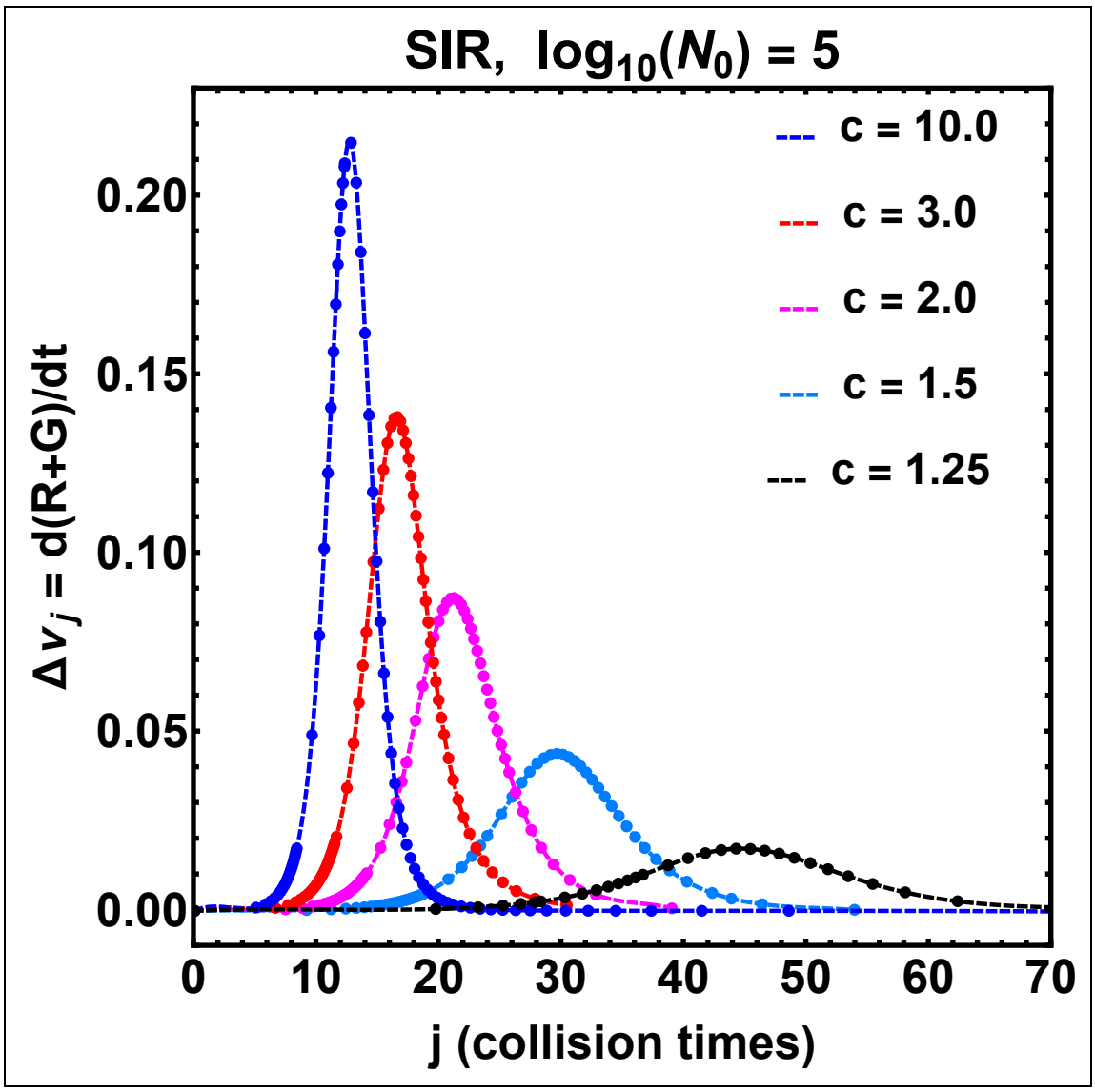

Fig. 33.— The growth rate, $d(R+G)/dt$, calculated from the SIR-model using Eq.(9.1a) with Eqs.(D3), (D5), and (D7) for values of $c = 1.25, 1.5, 2.0, 3.0$ and 10. These curves are analogous to the plots of $\Delta\nu_j$ in Figure 7, showing the weakening, widening, and shifting of the growth-rate profiles as c approaches 1.

that $R + G = 1 - 1/c$. Then, neglecting ϵ in Eq.(D3) for B, we find that $e^{-cG} = 1/c$ which gives $G = ln(c)/c$ and $R = 1 - (1/c) - \{ln(c)/c\}$. Summarizing these results, in the

SIR-model, R has its maximum value when

$$B = \frac{1}{c}, \tag{D9a}$$

$$R + G = 1 - \frac{1}{c}, \tag{D9b}$$

$$G = \frac{ln(c)}{c}, \tag{D9c}$$

$$R = 1 - \frac{1}{c} - \frac{ln(c)}{c}. \tag{D9d}$$

These equations are identical to the classic herd equations for the SIR-model if we regard c to be the epidemiological parameter, R_0. In particular, if $R_0 = c = 1.5$, then $B = 0.67$, $R + G = 0.33$, $G = 0.27$, and $R = 0.06$ when $dR/dt = 0$. If R_0 and c are larger, say $R_0 = c = 3$, then $B = 0.33$, $R + G = 0.67$, $G = 0.37$, and $R = 0.30$ when $dR/dt = 0$.

Figure 34 provides a graphic comparison of the herd-immunity threshold for the SIR-

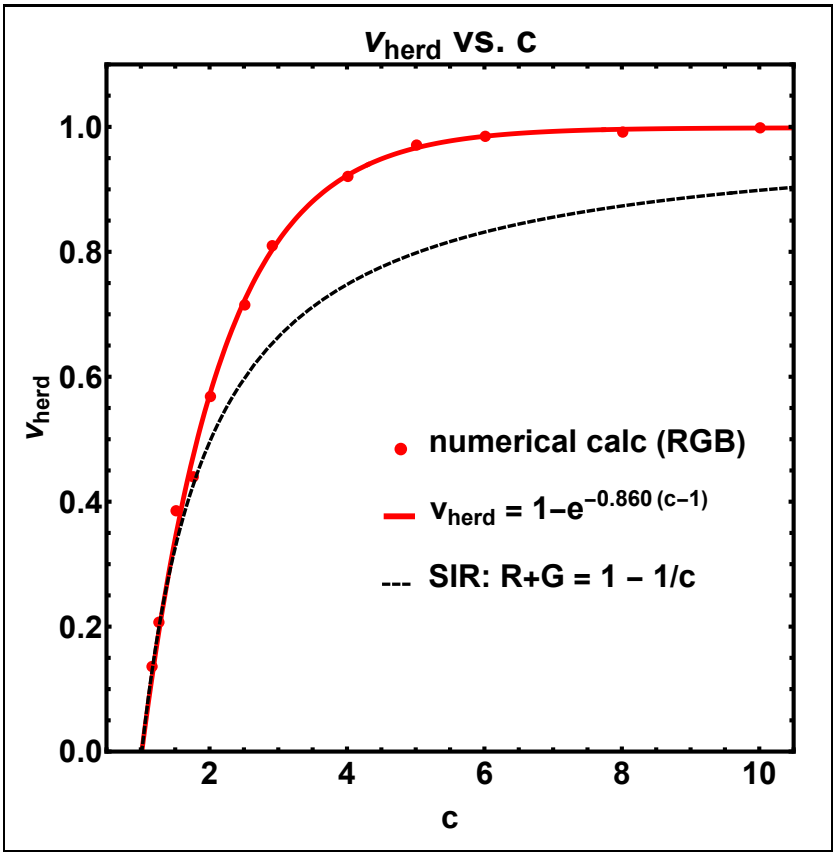

Fig. 34.— The value of ν_{herd}, evaluated numerically for the RGB-model and from $R + G = 1 - 1/c$ for the SIR-model. In each case, ν_{herd} is the total number of infections $(R+G)$ when the number of contagious red molecules reaches its peak.

model and the RGB-model. The red curve is the root-mean-square best fit to the computed

data points for the RGB-model using the exponential given by

$$\nu_{herd} \;=\; 1 \;-\; e^{-0.860(c-1)} \tag{D10}$$

and the dashed black curve is the SIR-expression for $R+G = 1-1/c$ given by Eq.(D9b). Although the two models show the same (normalized) number of infections for $c < 1.5$, the SIR-model shows a smaller number and a slower approach to 1 as c becomes larger.

In Figure 35, we show the difference, $j_{red} - j_{max}$, between the times that R and $R+G$ reach their respective peaks in the SIR-model. The j-values for each point were obtained

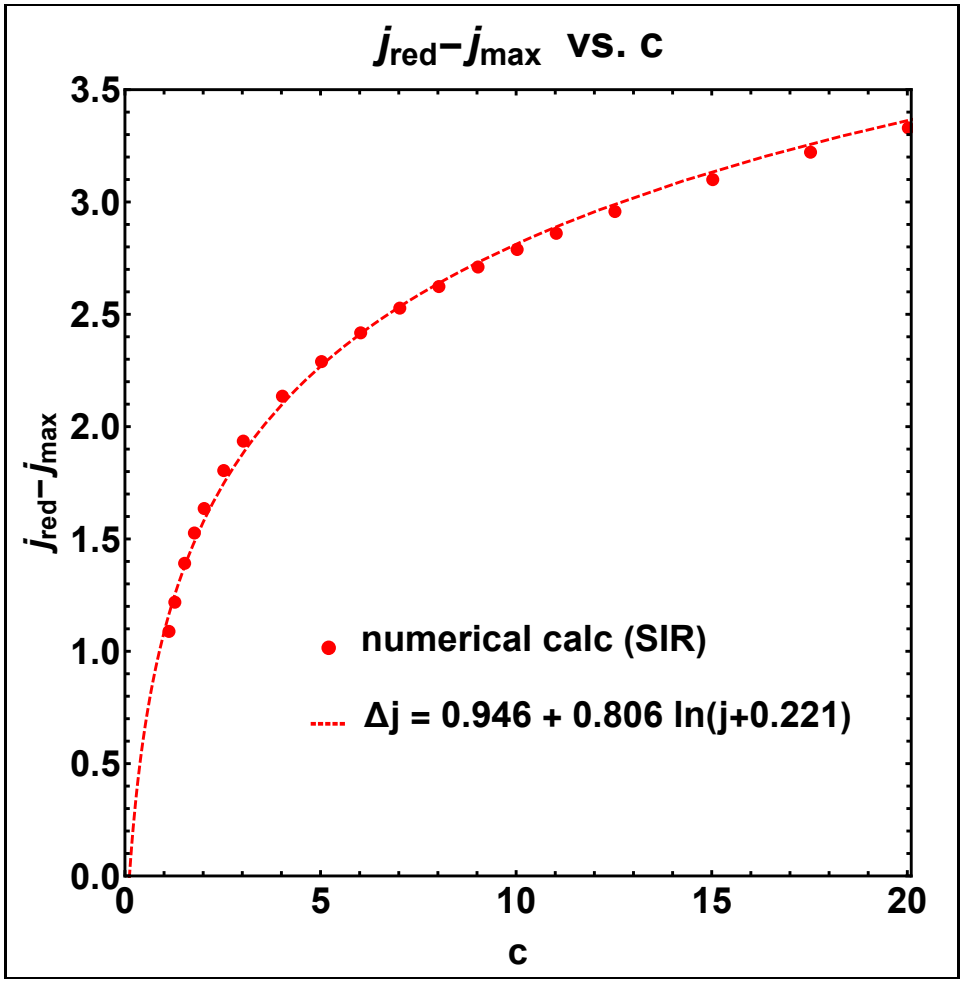

Fig. 35.— The time lag, $j_{red} - j_{max}$, between the location of the peak for infected molecules, $R+G$, and the peak for contagious molecules, R, expressed in units of the average time between collisions, τ_s, and plotted as a function of the contagious parameter, c, for the SIR-model.

by evaluating the integral in Eq.(D7) for two different upper limits; for j_{red}, we used the expression, $ln(c)/c$, given in Eq.(D9c); for j_{max}, we used the root of the equation $G+2e^{-cG} = (c+1)/c$, which results from combining $d^2(R+G)/dt^2 \sim G'' - c(G')^2 = 0$ with the expression

for $G' = \gamma(1 - G - e^{-cG})$ derived from Eqs.(9.1c) and (D5). The dashed red curve is the root-mean-square best fit to the data points, as given by the equation

$$j_{red} \; - \; j_{max} \; = \; 0.946 \; = \; 0.806 \; ln(c + 0.221). \tag{D11}$$

The data points used for this fit extend outward to $c = 100$, which has a j-value of $j_{red} - j_{max} = 4.732$. So although the lag does not increase rapidly with c, it nevertheless has not reached a limit by $c = 100$ (or even $c = 1000$, when the lag is 6.934). Like the corresponding lag determined for the RGB-model in Figure 19, this best-fit logarithmic expression for the SIR-model has some curvature and increases very slowly with c. By comparison, 'purely antisymmetric' models, like the hyperbolic-tangent and square-root models described in the footnote to section (5.2) and in section (6.2), have lags that increase linearly as $(c - 1)/2$ before they saturate.

Finally, we note that the epidemiological parameter, R_0, is often defined as the number of secondary infections that are produced in a homogeneous population of susceptible blue molecules by a single contagious red molecule before it loses its contagiousness and turns green. In the SIR-model, this works out to $R_0 = \beta/\gamma$, which we have called c. However, in the RBG-model, the number is $2^c - 1$ because each infected molecule goes on to infect other blue molecules before the original red molecule suddenly loses its contagiousness. These secondary infections continue to infect blue molecules without the exponential weakening that gives the average decay time, $1/\gamma$, in the SIR-model. As mentioned above, we can recover the classic SIR expression

$$\nu_{herd} \; = \; 1 \; - \; \frac{1}{c} \; = \; 1 \; - \; \frac{1}{R_0} \tag{D12}$$

by substituting $c = R_0$ in Eq.(D9b). However, the substitution of $2^c - 1 = R_0$ in RGB Eq.(6.4) gives

$$\nu_{herd} \; = \; 1 \; - \; e^{-0.860(c-1)} \; = \; 1 \; - \; (\frac{2}{1 + R_0})^{1.24}. \tag{D13}$$

These two equations give the value of ν_{herd} (or $R + G$) when the number of red molecules reaches its peak. In Figure 34 we saw that these thresholds occur for different values of c. Now, comparing Eqs. (D12) and (D13), we can also see that the thresholds occur for different values of R_0 (except when $c = R_0 = 1$, so that $\nu_{herd} = 0$ and there is no epidemic).

Acknowledgements

I am grateful to Dr. Stephen Kennedy (Mathematical Association of America Press) and Dr. Pete Riley (Predictive Science, Inc) for helpful comments related to the publication of this manuscript.

References

Anderson, R. M. and May, R. M. (1979), 'Population Biology of Infectious Diseases: Part I', Nature 280, 361-367.

Berra, Yogi, (1998), 'The Yogi Book', Workman Publishing Co., New York

Delamater, P. L., Street, E. J., Leslie, T. F., Yang, Y., & Jacobsen, K. H., (2019). Complexity of the Basic Reproduction Number (R0). Emerging Infectious Diseases, 25(1), 1-4. https://dx.doi.org/10.3201/eid2501.171901.

Fine, Paul; Eames, Ken; and Heymann, David L., (2011), ' 'Herd Immunity': A Rough Guide', Invited article on vaccines, Stanley Plotkin (ed.) in Clinical Infectious Diseases 2011; 52 (7):911-916, Oxford Univ. Press.

Huppert, A. and G. Katriel, 2013, 'Mathematical modeling and prediction in infectious disease epidemiology', Clinical Microbiology and Infection, Vol. 19, Issue 11, 999-1005.

Jones, D. S. and Sleeman, B. D. (1983), Ch 14 in 'Differential Equations and Mathematical Biology', Allen & Unwin, London.

Kermack, W. O. and McKendrick, A. G., 1927, 'A Contribution to the Mathematical Theory of Epidemics.' Proc. Roy. Soc. Lond. A 115, 700-721, 1927.

Leighton, R. B., 1964, 'Transport of Magnetic Fields on the Sun', Astrophys. J. 140, 1547.

Ying Liu, Albert A Gayle, Annelies Wilder-Smith, Joacim Rocklov, 2020, The reproductive number of COVID-19 is higher compared to SARS coronavirus, Journal of Travel Medicine, Volume 27, Issue 2, March 2020, taaa021, https://doi.org/10.1093/jtm/taaa021

Sanche, S., Lin, Y., Xu, C., Romero-Severson, E., Hengartner, N., & Ke, R., 2020. High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2. Emerging Infectious Diseases, 26(7), 1470-1477. https://dx.doi.org/10.3201/eid2607.200282.

Smith, D. A. and Moore, L. C., (1996), in 'Calculus: Modeling and Application', D. C. Heath and Co., Lexington MA.

Wang, Y. -M. and Sheeley, N. R., Jr., 1994, 'The Rotation of Photospheric Magnetic Fields: A Random Walk Transport Model', Astrophys. J. 430, 399-412.

Weisstein, Eric W., 2004, 'Kermack-McKendrick Model.' From MathWorld–A Wolfram Web Resource. https://mathworld.wolfram.com/Kermack-McKendrickModel.html