Multishapes and Packability in Statistical Geometry

One of the interesting questions about statistical geometry [1-3] fractals is "How does the number of needed trials vary with the shape? Toward this end I have constructed code which fractalizes "multishapes" -- blobs with varying degrees of roughness and thus differing "packability".

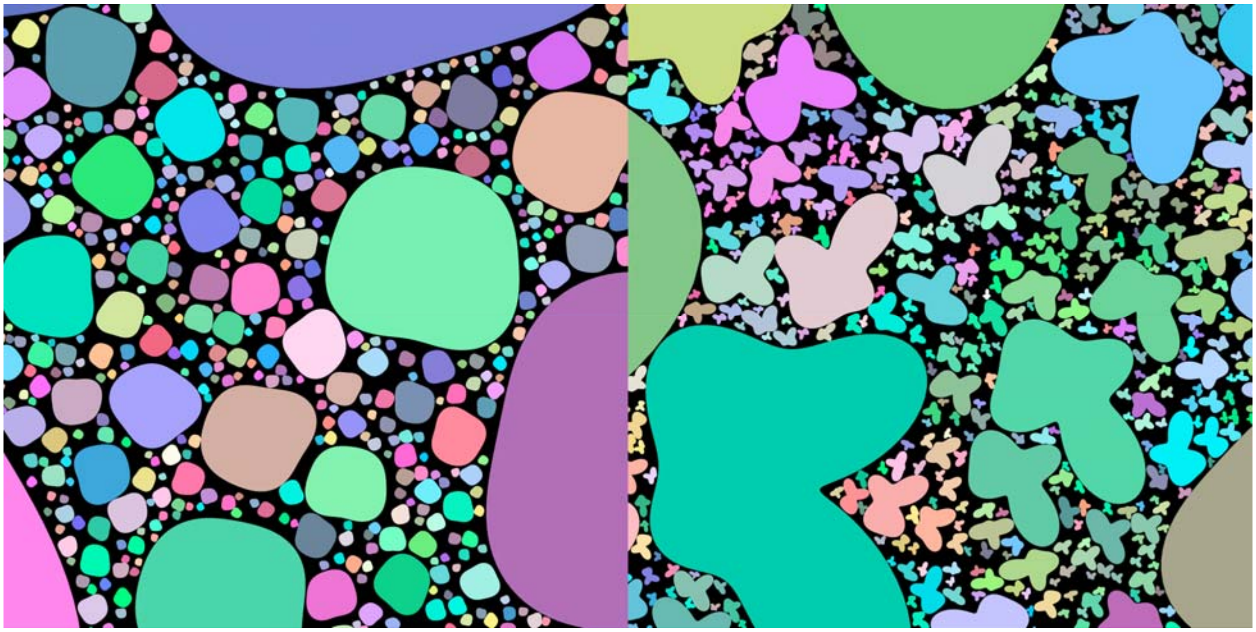


Fig. 1. Some examples of placed multishapes. In both cases c = 1.2 and N = 1. In the left-hand picture δ = .05 while on the right δ = .25. Note the correlation of neighboring shapes on the right-hand side (shown by the "islands" of similar colors/shapes). Color is a continuous function of shape (see text).

The basic algorithm is described in Appendix A and in the references.

1. The Multishape Defined.

In this study the shape was defined in polar coordinates relative to the randomly-chosen origin point x,y. The equation for a multishape is

$$r(\theta) = R(1+\delta[\cos(2\theta+\phi_1)+\cos(3\theta+\phi_2)+\cos(4\theta+\phi_3)]) \qquad \text{Eq. (1)}$$

One might expect a $cos(\theta+\phi_0)$ term but it is omitted as it simply results in another circle of larger area with its center shifted. The phase angles ϕ are random variables uniformly distributed over the interval from 0 to 2π. If one wishes to avoid negative r values, $\delta < 1/3$.

The resulting shapes go from circles when $\delta = 0$ to complicated many-lobed things as δ increases. Thus we have a shape whose "roughness" can be continuously varied. This can be seen in Fig. 1. The most prominent features for large δ are the sharp "dents" and the broad lobes.

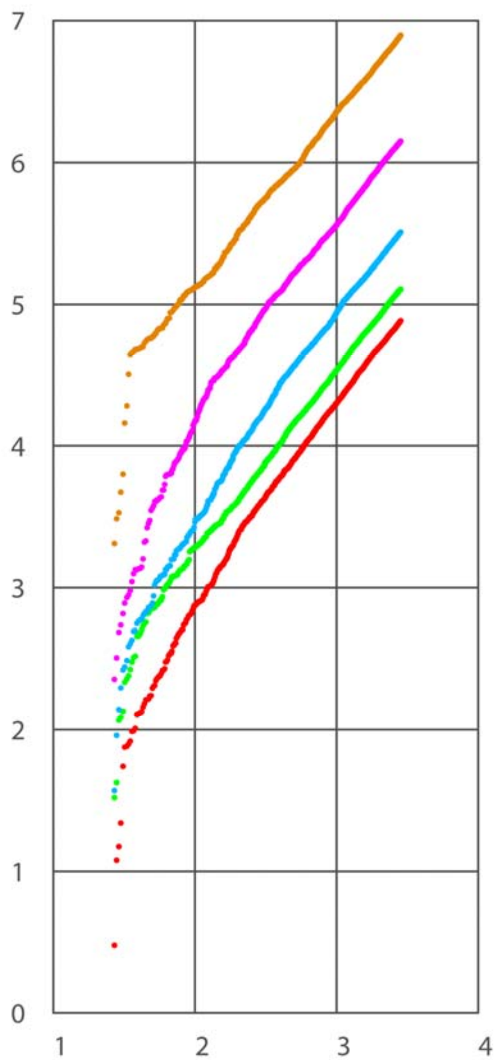
In the statistical geometry algorithm one must know the area of the shape. A fairly straightforward exercise in calculus shows that

$$A = \pi R^2 (1 + 1.5\delta) \qquad \text{Eq. (2)}$$

when 3 terms are used. A more general formula is given in Appendix C.

<u>The color scheme encodes the shape</u>. The shape is defined only by the three phase angles ϕ. The RGB color system has 3 dimensions, and the strength of the red color is made proportional to ϕ_1, etc. so that each color uniquely encodes a shape. This was done to see if there is correlation in the shapes. Each new trial is made with a new set of (ϕ_1, ϕ_2, ϕ_3) with random values uniformly distributed between 0 and 2π.

2. The Number of Trials, Effect of δ.



The total number of trials (vertical) versus the number of placed shapes n (horizontal) in log-log coordinates. The numbers next to the lines show the corresponding power of 10, i.e., the "2" line corresponds to a value of 10 squared, or 100.

c=1.20 N=1 Fill factor 81.7% in all cases.

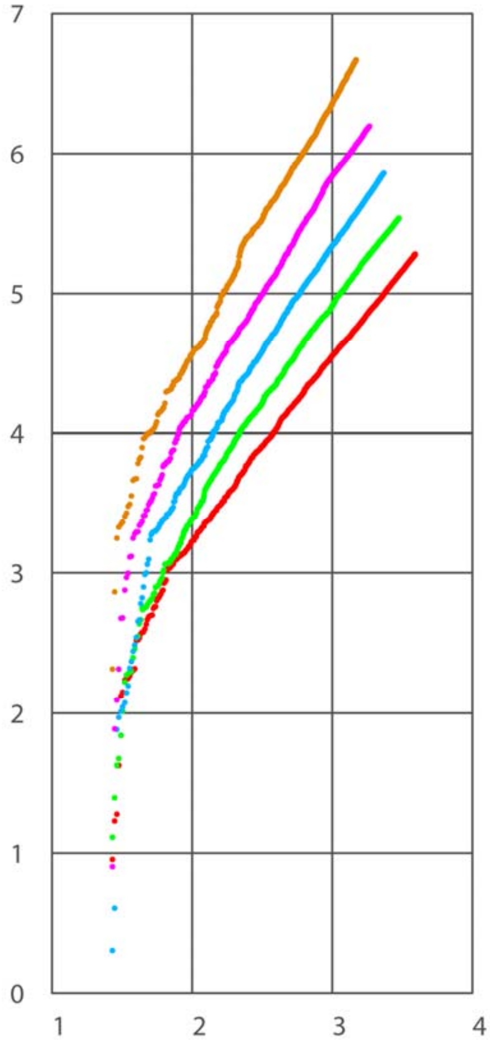
brown - delta=.25
magenta - delta=.20
blue - delta=.15
green - delta=.10
red - delta = .05

It can be seen that there is quite a bit of irregularity in the process for small n (left side of the graph), but for large n values the curves become parallel lines with the same slope. This indicates that the cumulative number of trials versus n follows a power law with the same exponent for all values of delta. The log-log slope is the exponent in this power law, and is called f. The values of f found by least-squares regression run from about 1.27 to 1.30 for the different curves.

The number of trials needed increases very sharply with increasing delta.

Fig. 2. Log-log plots of the cumulative number of trials needed to place n shapes. It can be seen that the data follows a straight line for large n in all cases.

3. The Effect of the Parameter *c*.



the total number of trials (vertical) versus the number of placed shapes n (horizontal) in log-log coordinates with variable parameter c. The numbers next to the lines show the corresponding power of 10. In call cases the process was stopped when the nominal radius of the last shape was .007 times the first radius.

delta=.10 N=1

brown - c=1.36
magenta - c=1.32
blue - c=1.28
green - c=1.24
red - c=1.20

the exponent f of the cumulative trials versus number n of placed shapes varies here -- the straight lines at the right of the graph are not parallel. The estimates of the exponent f in the power law are:

brown - f=1.73
magenta - f=1.52
blue - f=1.44
green - f=1.32
red - f=1.25

The variation in the K factor in Eq. (5) is much less pronounced here.

Fig. 3. Log-log plots of the cumulative number of trials needed to place *n* shapes with constant δ and variable *c*. While variation in δ does not change the power law exponent *f*, changes in *c* affect it strongly. A crude extrapolation of the straight-line segments with a ruler shows that they approximately converge at $n = 1$ (where $\log_{10}(n) = 0$).

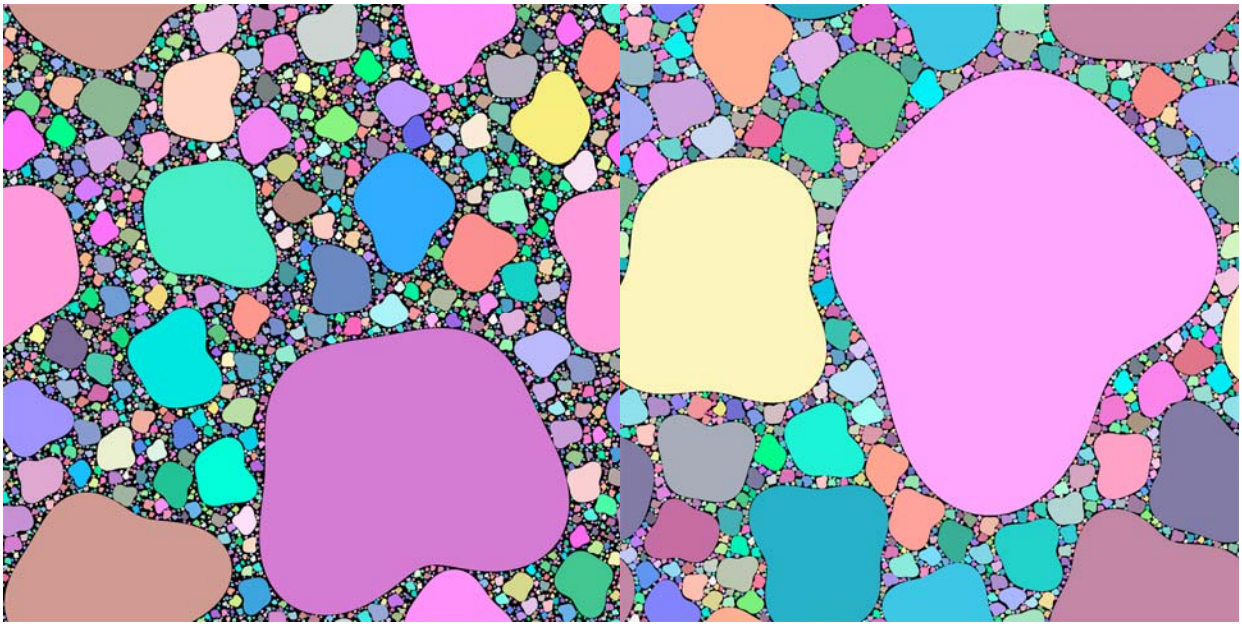


Fig. 4. Two examples of the fractal patterns corresponding to the cases in Fig. 3. On the left c = 1.20 while on the right c = 1.36. In neither case does it appear that changing c makes an observable difference in the correlation, which appears to the eye to be random in both instances. Apparently only δ affects correlation. The packing (fill) is much tighter with c = 1.36, which agrees with what is seen in the simple cases of circles and squares. Periodic boundary conditions.

4. Amoebas -- a Difficult Case.

It was decided to look at a more complicated shape by using a different set of sinusoids. The only change was to use the 3rd, 5th, 7th harmonics instead of 2nd, 3rd, 4th (see Eq. (3)). Note that 3,5,7 are all prime numbers and have no common divisor. This produces a much more varied and hard-to-fit set of shapes which I have called "amoebas".

$$r(\theta) = R(1 + \delta[\cos(3\theta + \phi_1) + \cos(5\theta + \phi_2) + \cos(7\theta + \phi_3)]) \qquad \text{Eq. (3)}$$

In the example c = 1.32, N = 1, and δ = .15.

The resulting fractal is shown in Fig. 5, and a plot of the data in Fig. 6. This shape has very low packability. A reasonable definition of packability as a number remains to be found.

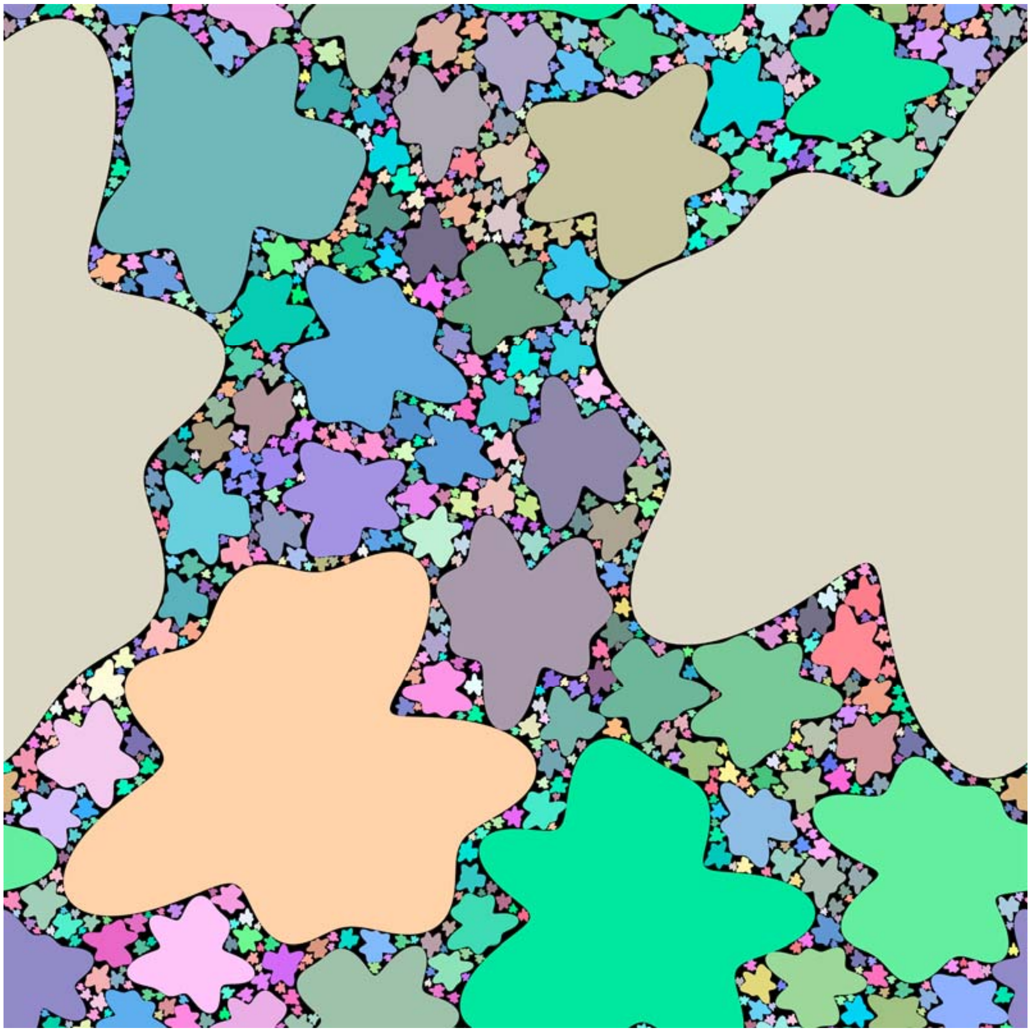


Fig. 5. "Amoebas" created by using higher-frequency harmonics of orders 3,5,7 (see Eq. (3)). There does appear to be a degree of correlation here -- substantially more so than with the simpler multishapes with the same δ. The inventive viewer can see turtles, trees, ponds, snowmen, ink blots, catsup splats, ghosts, … .

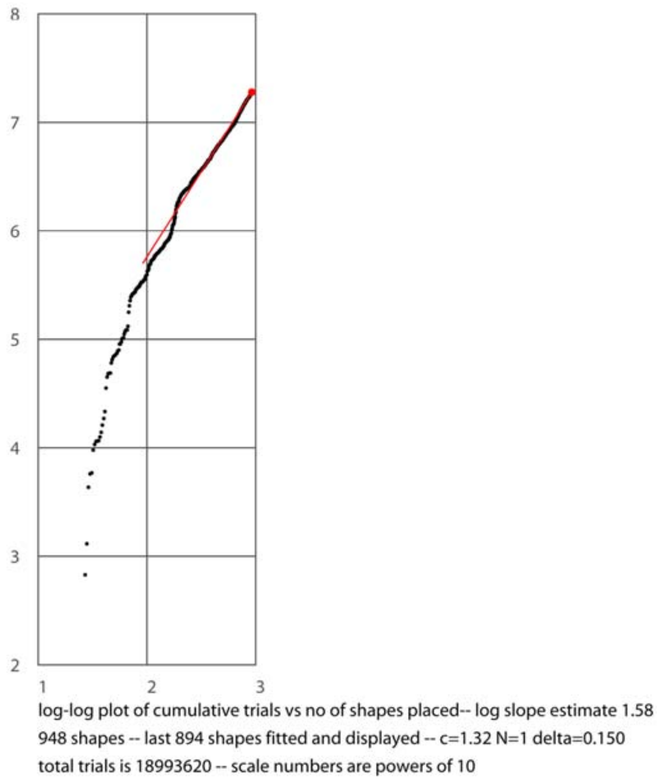


Fig. 6. Data plot for the "amoebas". Despite the very slow fractalization it appears that the $n_{cum}(n)$ curve assumes a power law with the modest exponent of f = 1.58 -- not far from what was seen in Fig. 3 for c = 1.32. Fractalization is slow not because of a high f but because of a high K (see Eq. (5)). The "oscillations" around the main trend of the data are often seen for hard-to-fractalize cases.

5. Gears.

This is a simple shape to describe and provides a nice test of packability (see Fig. 7).

$$r(\theta) = R(1+\delta[\cos(7\theta+\phi_1)]) \qquad \text{Eq. (4)}$$

The two key parameters for adjustment in packability studies are the number of teeth (chosen here to be the prime number 7) and the fractal exponent c.

With the relatively high c value (1.29) used here one must have partial "meshing" of the gears to achieve placement, and while it runs very slowly the exponent f is not particularly large.

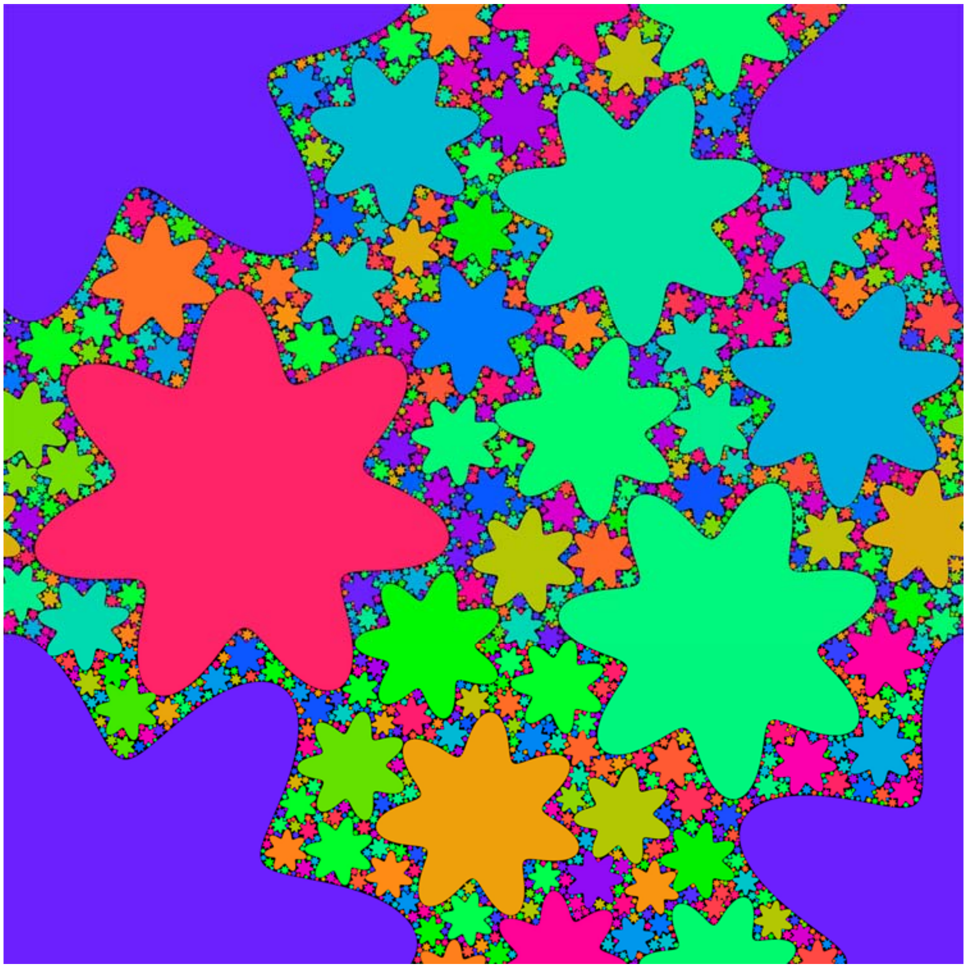


Fig. 7. Fractalized "gears". $c = 1.29$, $N = 1$, 3693 shapes, 92.1% fill. 120 927 421 trials (for an average of 32745 trials per placement). $f \cong 1.65$. The color is a continuous function of the rotation angle; there is no obvious correlation.

6. Discussion and Conclusions.

- It is evident that similarity of shapes is not a requirement for fractalization. For the multishapes no two shapes are congruent with each other.

- The n_{cum} versus n curve[1] follows a power law for large n in all cases.

$$n_{cum}(n) \cong Kn^{f} \qquad \text{Eq. (5)}$$

[1] I have worked with the cumulative data rather than the data for average number of trials needed to place the n-th shape. The biggest advantage of this is that it has some built-in averaging and is far less noisy. One can find the average number of trials needed for a given n by taking the derivative of $n_{cum}(n)$ with respect to n. This results in a power law with exponent $f-1$.

- This data supports the claim that for any number *n* of shapes to be placed the number *m* of required trials is always finite and predictable, i.e., the algorithm never halts[2].

- The exponent *f* in this power law is the same for all values of δ (within statistical error) with a fixed *c*, but varies substantially as *c* changes.

- The factor *K* in the power law of Eq. (5) increases very rapidly as δ increases (with fixed *c*). All of the increase in the number of trials needed for *n* placements can be ascribed to changes in *K*.

- The run time penalty for a hard-to-fractalize shape is quite large. *K* increases 100-fold between δ = .05 and δ = .25 (with *c* = 1.20).

- The process (algorithm) appears to have a "transient" region for low *n*, and settle down to a predictable "steady state" power-law condition for larger *n*. There are thus two regions of behavior.

- There is no sign that there is a limit for δ beyond which the multishapes can no longer be fractalized.

- The algorithm appears to have two regions of operation. For low *n* there is a very steep rise in the number of trials per placement with increasing *n*. $n_{cum}(n)$ then goes over to a power law for large *n* (Eq. (5)). These regions can be thought of as "initial transient" and a "steady state" regimes.

The strongest conjecture that one can make is that "*The statistical geometry algorithm can fractalize any shape*". If true, this is a surprising thing and suggests that fractalizability is a "property of two-dimensional space". Another conjecture is that *it works for any sequence of shapes so long as the area law is obeyed*.

Shapes which cannot be fractalized have yet to be found despite substantial searching by the author and others. Some shapes fractalize *very* slowly. Hard-to-fractalize shapes seem to work better at lower *c* values.

The algorithm shows fairly strong correlation of the multishapes for large δ. This can be seen in Fig. 1. On the left-hand side there is very little mutual correlation, which is shown by the fact that the colors are quite random. On the right-hand side where the packability is low and many trials are needed the

[2] There are two kinds of numbers that one can think of in considering whether the algorithm stops. In the practical execution of the algorithm one uses finite-precision floating point numbers. In the mathematical world one would think of infinite-precision numbers. With finite precision the algorithm will eventually stop. The best assessment one can make at this time is that with "mathematician's" numbers the process will continue indefinitely in a fractal "self-similar" way. It is noted, however, that computational experiments fall far short of a rigorous proof.

correlation is fairly strong. There are islands of similar colors showing that the process[3] must "choose" similar shapes within a given region to achieve placement.

It would be interesting to find a mathematical description of the correlation phenomena in these fractals so that a number or function would describe the average correlation, but this remains to be done.

7. References.

[1] Statistical geometry article on the web site john-art.com.

[2] John Shier, "Hyperseeing", Summer 2011 issue, pp. 131-140, published by ISAMA. Available by download at the web site john-art.com.

[3] "Statistical Geometry", John Shier, July 2011. A colorful self-published fractal art picture book available at lulu.com.

[4] "The Fractal Geometry of Nature", Benoit Mandelbrot, 1977.

John Shier

August 2011

Appendix A. The Statistical Geometry Algorithm.

It has been found [1]-[3] that it is possible to create fractal patterns of a wide variety of geometric shapes by the following algorithm:

1. Create a sequence of areas A_i equal to $\frac{1}{N^c}, \frac{1}{(N+1)^c}, \frac{1}{(N+2)^c}, \frac{1}{(N+3)^c}, \ldots$. Choose an area (square, rectangle, circle, …) A to be filled.

2. Sum the areas A_i to infinity, using the Hurwitz zeta[4] function

$$\zeta(c,N) = \sum_{i=0}^{\infty} \frac{1}{(N+i)^c}$$

3. Define a new set of areas S_i by $S_i = \frac{A}{\zeta(c,N)(N+i)^c}$. It will be seen that the sum of all these redefined areas is just A.

---

[3] The author is inclined to see parallels between this "random search" process and the widely studied "random walk" processes. Both processes involve randomness constrained by previous events, i.e., constrained randomness. Both processes are susceptible to many variations.

[4] The definitions of the Riemann and Hurwitz zeta functions can be found in Wikipedia.

4. Let $i = 0$. Place a shape having area S_0 in the area A at a random position x,y such that it falls entirely within area A. Increment i. This is the "initial placement".

5. Place a shape having area S_i entirely within A at a random position x,y such that it falls entirely within A. If this shape overlaps with any previously-placed shape repeat step 5. This is a "trial".

6. If this shape does not overlap with any previously-placed shape put x,y and the shape dimensions in the "placed shapes" data base, increment i, and go to step 5. This is a "placement".

7. Stop when i reaches a set number, percentage filled area reaches a set value, or other.

One will note that the dimensions of the shapes are nowhere specified. They are calculated from the areas. A very wide variety of shapes have been found to be "fractalizable" in this way.

The parameters c and N can have a variety of values. The parameter c is often in the range 1.2-1.4 with a largest usable value around 1.51. N can be 1 or larger, and need not be an integer.

By construction the result is a space-filling random fractal -- if the process never halts. Available evidence [1]-[3] says that it does not. The power law area sequence ensures that it has the fractal "statistical self-similarity" (scale-free) property. And the random search ensures that no two circles will ever touch, so that the "gasket" is a single continuous object.

Appendix B. Data Fitting.

The raw data were converted to $\log_{10}$ values. These values were then fitted by least squares adjustment to a straight line with the points weighted as the y value (the cumulative number of trials required). Note that the weighting was done with the n_{cum} value, not its logarithm. This weighting had the effect of emphasizing the right-hand side of the curves in Fig. 2 so that the slopes (exponents) found relate to the "steady state" part of the data (large n).

Appendix C. The Area of a Multishape.

Let the polar equation $r(\theta)$ for a shape be given by

$$r(\theta) = R(1 + \sum_{m=1}^{M} a_m \cos(m\theta + \phi_m)) \qquad \text{(C1)}$$

The area in polar coordinates is

$$A = \frac{1}{2}\int_0^{2\pi} r(\theta)^2 \, d\theta \qquad \text{(C2)}$$

If we insert our general formula (C1) into this it becomes

$$
\begin{aligned}
A &= \frac{R^2}{2}\int_0^{2\pi}(1+\sum_{m=1}^{M} a_m \cos(m\theta+\phi_m))^2\, d\theta \\
&= \frac{R^2}{2}\int_0^{2\pi}(1+2\sum_{m=1}^{M} a_m \cos(m\theta+\phi_m)+\left[\sum_{m=1}^{M} a_m \cos(m\theta+\phi_m)\right]^2)d\theta
\end{aligned} \qquad \text{(C3)}
$$

The first term in the integral is just 2π. The second term is zero because such a sinusoid is as often positive as negative and the + and – contributions cancel. The third term produces a big mess of squared and cross terms, but this simplifies greatly when one observes that the sinusoids are orthogonal[5] functions, hence only the square terms survive in the integration. Thus

$$
\begin{aligned}
A &= \frac{R^2}{2}(2\pi+\sum_{m=1}^{M} a_m^2 \pi) \\
&= \pi R^2\left(1+\frac{1}{2}\sum_{m=1}^{M} a_m^2\right)
\end{aligned} \qquad \text{(C4)}
$$

When all of the a_m are zero the formula goes over to the well-known equation for the area of a circle of radius R (as it must).

[5] The orthogonality property of sinusoidal "harmonics" is described in 100s of textbooks on mathematics, science, and engineering, usually under the heading of "Fourier series".