

Claremont Colleges Scholarship @ Claremont

All HMC Faculty Publications and Research

HMC Faculty Scholarship

1-1-2005

SRT Division Algorithms As Dynamical Systems

Mark McCann
Princeton University

Nicholas Pippenger
Harvey Mudd College

Recommended Citation

McCann M., and Pippenger N. "SRT Division Algorithms as Dynamical Systems", *SIAM Journal of Computing*, 34, 1279 (2005).

This Article is brought to you for free and open access by the HMC Faculty Scholarship at Scholarship @ Claremont. It has been accepted for inclusion in All HMC Faculty Publications and Research by an authorized administrator of Scholarship @ Claremont. For more information, please contact scholarship@cuc.claremont.edu.

SRT DIVISION ALGORITHMS AS DYNAMICAL SYSTEMS*

MARK MCCANN[†] AND NICHOLAS PIPPENGER[†]

Abstract. Sweeney–Robertson–Tocher (SRT) division, as it was discovered in the late 1950s, represented an important improvement in the speed of division algorithms for computers at the time. A variant of SRT division is still commonly implemented in computers today. Although some bounds on the performance of the original SRT division method were obtained, a great many questions remained unanswered. In this paper, the original version of SRT division is described as a dynamical system. This enables us to bring modern dynamical systems theory, a relatively new development in mathematics, to bear on an older problem. In doing so, we are able to show that SRT division is ergodic, and is even Bernoulli, for all real divisors and dividends. With the Bernoulli property, we are able to use entropy to prove that the natural extensions of SRT division are isomorphic by way of the Kolmogorov–Ornstein theorem. We demonstrate how our methods and results can be applied to a much larger class of division algorithms.

Key words. SRT division, ergodic, Bernoulli, dynamical systems, entropy

AMS subject classifications. 68W40, 37E05

DOI. 10.1137/S009753970444106X

1. Introduction. Since the discovery of the first radix-2 Sweeney–Robertson–Tocher (SRT) division algorithm, the use of the term “SRT division” has expanded to include a wide variety of higher radix nonrestoring division algorithms that are loosely based on the original. For example, there is the infamous implementation of a radix-4 SRT division algorithm in the first release of the Pentium CPU that has become widely known as the “Pentium Bug.” One major difference between this implementation of radix-4 SRT division and the original radix-2 SRT division is that the former produces a constant number of quotient bits per step, while the latter produces a variable number. Modern implementations of SRT division use carry-save adders to perform additions and subtractions in constant time. Earlier implementations, however, used carry-propagate adders with delays that grow with the word length. Therefore, the primary goal of the early investigators was to reduce the number of uses of the costly adder. In the late 1950s, Sweeney [3], Robertson [17], and Tocher [21] independently made the observation that whenever a partial remainder is in the range $(-\frac{1}{2}, \frac{1}{2})$, there will be one or more leading zeros that can be shifted through in a very short amount of time (usually one cycle) thereby reducing the use of the adder. Although the aforementioned have received most of the credit for the algorithm named after them, it can be argued that Nadler described an equivalent algorithm in a 1956 paper [13]. The description of higher radix SRT division, which is the basis for modern SRT division, is generally attributed to Atkins [1], but this is not the version of division that we will be concerned with in this paper.

*Received by the editors February 16, 2004; accepted for publication (in revised form) February 3, 2005; published electronically June 27, 2005. A preliminary version of this paper appears in [11].
<http://www.siam.org/journals/sicomp/34-6/44106.html>

[†]Department of Computer Science, Princeton University, 35 Olden Street, Princeton, NJ 08544 (mmccann@cs.princeton.edu, njp@cs.princeton.edu). The research of the first author was supported by an NSERC research grant. It was conducted while the author was at the University of British Columbia. The research of the second author was supported by an NSERC research grant and by a Canada Research Chair. It was conducted while the author was at the University of British Columbia.

Although what is considered to be “costly” for a division algorithm has changed, it is still interesting to study and important to understand the behavior of successive partial remainders on average for a given divisor. Surprisingly, some of the most basic questions that one might have concerning the behavior of partial remainders for even simple radix-2 SRT division have remained unanswered for over 40 years. The difficulty that early investigators experienced in answering such questions was mainly due to a lack of necessary mathematical tools and results. During that past 30 years, the field of “dynamical systems theory” or “ergodic theory” has come into existence in mathematics and has been greatly developed. In this paper we show how to apply some of what is now known in dynamical systems theory to the earliest version of SRT division. In doing so, we are able to prove several previously unknown properties for simple SRT division. The results are quite general and can be adapted to other division algorithms. We view the value of these results as lying in the establishment of a connection between a well-developed area of mathematics and digital division, rather than in any practical consequences for division algorithms. For the remainder of this paper, the term SRT division will refer to the original algorithm unless otherwise stated.

The SRT division algorithms analyzed by Freiman [5] and Shively [20] are the same, but the authors differ in what they take to be a step of the algorithm: Freiman defines a step to be the operations from one use of the adder to the next, while Shively defines it to be the operations from one normalizing shift (of a single place) to the next. The following definitions are consistent with Shively’s:

1. n represents the number of iterations performed in the algorithm.
2. p_0 is the dividend (or initial partial remainder) normalized so that $p_0 \in [\frac{1}{2}, 1)$.
3. $p_i \in (-1, 1)$, $i \in \mathbb{N}$, is the partial remainder after the i th step.
4. D is the divisor normalized to $[\frac{1}{2}, 1)$.
5. $q_i \in \{-1, 0, 1\}$ ($i \in \{0, \dots, n-1\}$) is the quotient digit generated by the i th step.
6. $Q_n = \sum_{i=0}^{n-1} \frac{q_i}{2^i}$ is the “rounded off” quotient generated after n steps of the algorithm.

Given the above definitions, after n steps of the division algorithm, we would like it to be true that

$$p_0 = DQ_n + \varepsilon(n),$$

where $\varepsilon(n)$ is a term that goes to zero as n goes to infinity.

A recurrence relation for the SRT division algorithm can be stated as

$$(p_{i+1}, q_i) = \begin{cases} (2p_i, 0) & : |p_i| < \frac{1}{2}, \\ (2(p_i - D), 1) & : |p_i| \geq \frac{1}{2} \text{ and } p_i \geq 0, \\ (2(p_i + D), -1) & : |p_i| \geq \frac{1}{2} \text{ and } p_i < 0. \end{cases}$$

By observing that

$$p_{i+1} = \begin{cases} 2(p_i - (0)D) & : |p_i| < \frac{1}{2}, \\ 2(p_i - (1)D) & : |p_i| \geq \frac{1}{2} \text{ and } p_i \geq 0, \\ 2(p_i - (-1)D) & : |p_i| \geq \frac{1}{2} \text{ and } p_i < 0, \end{cases}$$

we can rewrite the definition of p_{i+1} as

$$p_{i+1} = 2(p_i - q_i D).$$

After n steps have been completed, we have

$$p_n = 2^n p_0 - 2^n q_0 D - 2^{n-1} q_1 D - \dots - 2^1 q_{n-1} D,$$

and then after dividing by 2^n and solving for p_0 we find that

$$\begin{aligned} p_0 &= \frac{p_n}{2^n} + \frac{q_0 D}{2^0} + \frac{q_1 D}{2^1} + \dots + \frac{q_{n-1} D}{2^{n-1}} \\ &= D \sum_{i=0}^{n-1} \frac{q_i}{2^i} + \frac{p_n}{2^n} = D Q_n + \frac{p_n}{2^n}. \end{aligned}$$

Now let $\varepsilon(n) = p_n/2^n$ and let $Q^* = \lim_{n \rightarrow \infty} Q_n$. Since $|p_n| < 1$, in the limit as n goes to infinity,

$$p_0 = D Q^*.$$

The generated quotient bits ($-1, 0, +1$ valued) are not in a standard binary representation, but it is a simple matter to convert the answer back to standard binary without using any expensive operations. Table 1 shows an example of using the SRT division algorithm to divide 0.67 by 0.75. The steps that produce nonzero quotient bits have been shown. In this example, after six uses of the adder, the quotient (0.893) has been determined to four digits of precision.

TABLE 1
SRT division where $p_0 = 0.67$ and $D = 0.75$.

$p_0 = 0.67$	$= 0.67$	$q_0 = 1$	$Q_0 = 1$
$p_1 = 2(0.67 - D)$	$= -0.16$	$q_3 = -1$	$Q_3 = 0.875$
$p_4 = 2(2^2(-0.16) + D)$	$= 0.22$	$q_6 = 1$	$Q_6 = 0.890625$
$p_7 = 2(2^2(0.22) - D)$	$= 0.26$	$q_8 = 1$	$Q_8 = 0.89453125$
$p_9 = 2(2^1(0.26) - D)$	$= -0.46$	$q_{10} = -1$	$Q_{10} = 0.8935546875$
$p_{11} = 2(2^1(-0.46) + D)$	$= -0.34$	$q_{12} = -1$	$Q_{12} \doteq 0.8933105469$
$p_{13} = 2(2^1(-0.34) + D)$	$= 0.14$		

Now, with this simple system of division in hand, we might want to ask certain questions about its performance. For example, we could ask, “How many bits of precision are generated per iteration of the algorithm on average?” To answer this question, we must look at the magnitude of $|Q^* - Q_n| = |p_n/2^n|$. The number of bits of precision on the n th step is then $n - \log_2 p_n$. In the worst case, p_n is close to 1, and therefore we get at least one bit of precision per iteration of the algorithm, regardless of the values of D or p_0 . Of course, a designer of actual floating-point hardware probably wants to know the expected performance based on the expected values of p_n . To answer the many variants of this type of question, it is clear that we must know something about the distribution of partial remainders over time. The remainder of this paper is devoted to extending what is known about the answer to this type of question as it relates to SRT division and its variants.

2. SRT division as a dynamical system. The example in Table 1 makes it clear that keeping track of the signs of successive partial remainders is irrelevant in determining how many times the adder will be used for a particular calculation. For this reason, we need only consider the magnitudes of successive partial remainders. We now give a reformulation of SRT division that will allow us to look at division as a dynamical system.

DEFINITION 1 (SRT division transformation). For $D \in [\frac{1}{2}, 1)$, we define the function $T_D : [0, 1) \rightarrow [0, 1)$ as

$$T_D(x) = \begin{cases} 2x & : 0 \leq x < \frac{1}{2}, \\ 2(D - x) & : \frac{1}{2} \leq x < D, \\ 2(x - D) & : D \leq x < 1. \end{cases}$$

This transformation of the unit interval represents the successive partial remainders that arise as SRT division is carried out by a divisor D on a dividend x . D is normalized to $[\frac{1}{2}, 1)$. The dividend x is normalized to $[\frac{1}{2}, 1)$ initially, while each of the successive partial remainders $T_D^n(x)$ ($n \in \mathbb{N}$) subsequently ranges through $[0, 1)$.

By using the characteristic function for a set Δ defined as

$$1_\Delta(x) = \begin{cases} 1 & : x \in \Delta, \\ 0 & : x \notin \Delta, \end{cases}$$

we can rewrite T_D as

$$(1) \quad T_D(x) = 2x \cdot 1_{[0, \frac{1}{2})}(x) + 2(D - x) \cdot 1_{[\frac{1}{2}, D)}(x) + 2(x - D) \cdot 1_{[D, 1)}(x).$$

If we plot (1) on the unit interval, we obtain a very useful visualization of our transformation. Figure 1 shows the plot of $T_{0.75}(x)$ combined with a plot of the successive partial remainders that arise while dividing 0.67 by 0.75. This is the same system that was presented earlier in Table 1. Notice that a vertical line in the interval $[\frac{1}{2}, D)$ corresponds to a subsequent flip in the sign of the next partial remainder.

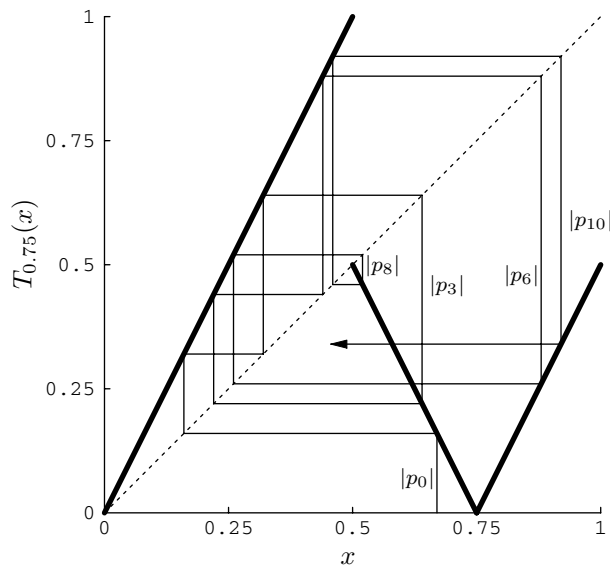


FIG. 1. Graphic representation of partial remainder magnitudes for $D = 0.75$ and $p_0 = 0.67$.

Figure 1 shows an example of following the trajectory of a single partial remainder for a particular divisor. In this figure, the heavy solid lines represent the transformation $T_{0.75}$, while the abscissa of the thin vertical lines represents successive partial

remainder magnitudes. After 10 applications of the $T_{0.75}$, there is not any obvious regular pattern, although we expect to see one eventually since the quotient is rational in this case.¹ Of course, most numbers are not rational and we can deduce that for most numbers, the transformation will never exhibit a repeating pattern. In Figures 2 and 3, we see that a very small change in the value of the initial partial remainder quickly produces large differences in the observed behavior of the subsequent partial remainders. As we show in the appendix, our system is actually chaotic, and therefore we will gain little understanding by studying the trajectories of individual partial remainders. The logical next step is to study the behavior of distributions of points over the whole interval.

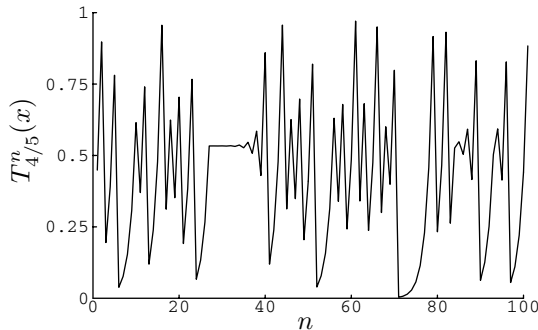


FIG. 2. The result of applying $T_{4/5}$ to $x = \frac{\pi}{7}$ one hundred times.

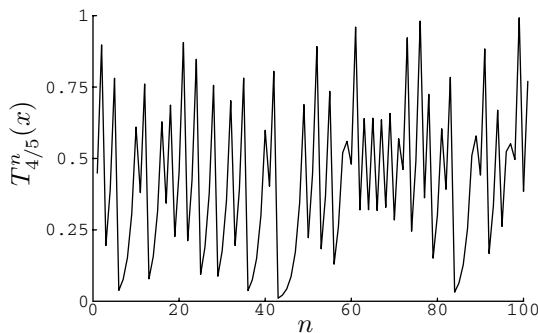


FIG. 3. The result of applying $T_{4/5}$ to $x = \frac{\pi}{7} + 0.00001$ one hundred times.

Understanding the behavior of ensembles of points under repeated transformation is in the realm of dynamical systems theory. For the remainder of this paper, we assume a certain amount of familiarity with the fundamentals of dynamical systems theory (or ergodic theory), which requires some basic understanding of measure theory. We will include a few helpful background material definitions as they are needed, but mostly we will provide references. A very good introduction to the study of chaotic systems is Lasota and Mackey's book [8]. For a more detailed introduction to ergodic theory (along with the necessary measure theory needed to understand this paper), Walters's book [22] and Petersen's book [16] are highly recommended.

¹With redundant representations, rational numbers can have aperiodic representations, though we do not expect this to happen.

DEFINITION 2 (probability space). *If \mathcal{B} is a σ -algebra on subsets of a set X and if m is a measure on \mathcal{B} , where $m(X) = 1$, then the triple (X, \mathcal{B}, m) is called a probability space. (See [22, pp. 3–9] and [8, pp. 19–31] for a good overview of basic measure theory and Lebesgue integration.)*

DEFINITION 3 (Perron–Frobenius operator). *For a probability space (X, \mathcal{B}, m) ,² the Perron–Frobenius operator $P : L^1 \rightarrow L^1$ associated with a nonsingular transformation $T : X \rightarrow X$ is defined by*

$$\int_B Pf(x) \, dm = \int_{T^{-1}(B)} f(x) \, dm \quad \text{for } B \in \mathcal{B} .$$

For a piecewise monotonic C^2 transformation³ T with n monotonic pieces, we can give an explicit formula for the Perron–Frobenius operator. Let $A = \{A_1, A_2, \dots, A_n\}$ be the partition of X which separates T into n pieces. For $i \in \{1, \dots, n\}$, let $t_i(x)$ represent the natural extension of the i th C^2 function $T(x)|_{A_i}$. The Perron–Frobenius operator for T is then

$$Pf(x) = \sum_{i=1}^n \left| \frac{d}{dx} t_i^{-1}(x) \right| f(t_i^{-1}(x)) \cdot 1_{t_i(A_i)}(x) .$$

In particular, for T_D (as in (1)),

(2)

$$Pf(x) = \frac{1}{2}f\left(\frac{1}{2}x\right) \cdot 1_{[0,1)}(x) + \frac{1}{2}f\left(D - \frac{1}{2}x\right) \cdot 1_{(0,2D-1]}(x) + \frac{1}{2}f\left(D + \frac{1}{2}x\right) \cdot 1_{[0,2-2D)}(x) .$$

With (2) we can show precisely what happens to an initial distribution of points (described by an integrable function) after they are repeatedly transformed under T_D . Figures 4 and 5 show what happens to two different initial distributions of points after five applications of the Perron–Frobenius operator associated with $T_{3/5}(x)$. By the fifth application, the distributions look remarkably similar. One might guess that they are both approaching the same final distribution. This situation is in marked contrast to chaotic behavior observed in Figures 2 and 3.

DEFINITION 4 (stationary distribution). *Let (X, \mathcal{B}, m) be a probability space, let P be the Perron–Frobenius operator associated with a nonsingular transformation $T : X \rightarrow X$, and let L^1 denote the L^1 space of (X, \mathcal{B}, m) . If $f \in L^1$ is such that $Pf = f$ a.e.,⁴ then f is called a stationary distribution of T .*

A practical use of the Perron–Frobenius operator is in deriving and verifying the equations of stationary distributions for given divisors. As an example of this, we verify the correctness of a previously known stationary distribution for $D \in [\frac{3}{4}, 1)$. An exact equation was first given by Freiman [5] and is restated by Shively [20] as

(3)

$$f(x) = \frac{1}{D}1_{[0,2D-1)}(x) + \frac{1}{2D}1_{[2D-1,1)}(x) .$$

This relation can be verified by applying the Perron–Frobenius operator as given in (2) to (3). Such exact equations are not known in general for all $D \in [\frac{1}{2}, \frac{3}{4})$. This issue is discussed further in the conclusion.

²For a probability space (X, \mathcal{B}, m) , the L^1 space of (X, \mathcal{B}, m) is the set of $f : X \rightarrow \mathbb{R}$ satisfying $\int_X |f(x)| \, dm < \infty$.

³ C^2 denotes the set of all functions with two continuous derivatives.

⁴a.e. indicates that the given relation holds *almost everywhere*, that is, everywhere except possibly on a set of measure zero.

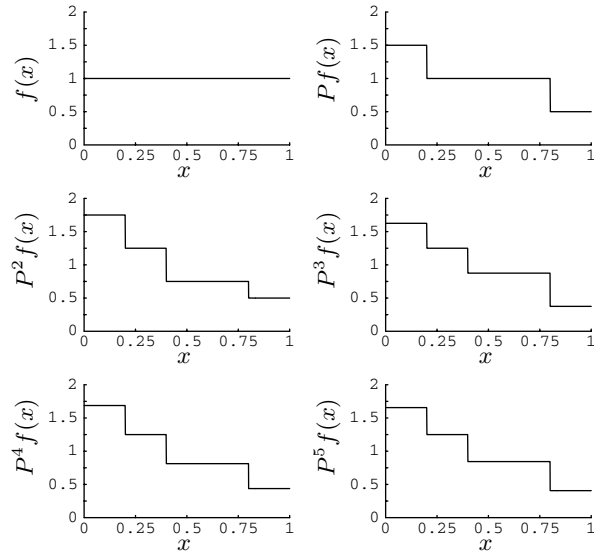


FIG. 4. The result of applying the Perron–Frobenius operator P associated with $T_{3/5}$ to $f(x) = 1$ six times.

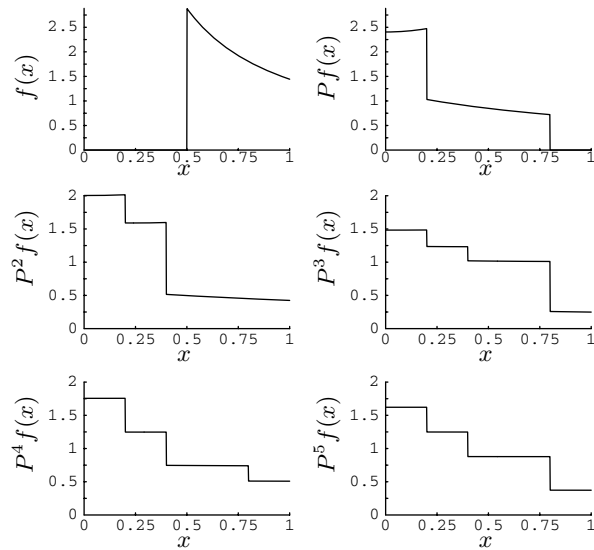


FIG. 5. The result of applying the Perron–Frobenius operator P associated with $T_{3/5}$ to $f(x) = (1/x \log_e 2) \cdot 1_{[\frac{1}{2}, 1)}(x)$ six times.

In the case of variable quotient-bits-per-cycle algorithms such as the original SRT division, one of the primary uses of a formula for the distribution of partial remainders is for calculating the *shift average* for a given divisor. (Note that higher shift averages are desirable.) The shift average is the average number of uses of the shift register (single shift or multiplication by two) between uses of the adder. Under the assumption

that a register shift is a much faster operation than using the adder, the shift average gives a useful characterization of the expected performance of our algorithm for a given divisor. With (3), we know the fraction of bits which require the use of the adder. To calculate the average number of zero bits generated between nonzero bits (bits requiring use of the adder), we take the reciprocal of the fraction of bits which do not require the adder. For $D \in [\frac{3}{4}, 1)$, the shift average function is

$$(4) \quad s(D) = \left(1 - \frac{1}{2D}\right)^{-1} = \frac{2D}{2D - 1}.$$

Unfortunately, since we have not proven that the stationary distributions from SRT division are unique, we have no way of knowing whether or not a shift average calculation in (4) is correct for all initial probability distributions. To prove that all stationary distributions are unique, we need to show that T_D is ergodic for all $D \in [\frac{1}{2}, 1)$.

DEFINITION 5 (ergodic; see [8]). *Let (X, \mathcal{B}, m) be a probability space and let a nonsingular transformation $T : X \rightarrow X$ be given. Then T is ergodic if for every set $B \in \mathcal{B}$ such that $T^{-1}(B) = B$, either $m(B) = 0$ or $m(X \setminus B) = 0$.*

Freiman [5] shows that T_D is ergodic for rational D , but we extend this result to real D . In the next section we show that all T_D are Bernoulli and it is known that having the Bernoulli property implies ergodicity.

3. Bernoulli property. Our central result, which we present in this section, is that the class of transformations of the interval that characterizes SRT division for all real divisors D has the property that each transformation T_D is Bernoulli. Although the basic concept of a Bernoulli shift (the things to which transformations having the Bernoulli property are isomorphic) is not difficult, a complete definition requires enough auxiliary concepts from measure theory (concepts not used anywhere else in this paper) that we refer the interested reader to [15, 16, 19, 22]. Neither an understanding of Bernoulli shifts nor a formal definition of what it means to be Bernoulli is required to follow the proofs in this section. Having said this, we should mention informally the connection between Bernoulli shifts and transformations having the Bernoulli property.

The transformation T_D is a noninvertible endomorphism of the unit interval. This means that from a given partial remainder we can predict all future partial remainders, but we cannot uniquely predict past partial remainders. There is a natural way (called the natural extension) to make our transformation invertible (an automorphism) on a larger space. Specifically, each noninvertible transformation T_D having the Bernoulli property has an extension to an automorphic transformation, isomorphic to a two-sided Bernoulli shift [16, pp. 13, 276]. From the way that entropy for a transformation is defined, the entropy for an automorphic Bernoulli transformation associated with a noninvertible Bernoulli transformation is the same as the entropy for the noninvertible Bernoulli transformation. By proving that all transformations T_D are Bernoulli, and by proving that the entropy of each T_D is the same, we will be able to conclude that the natural extensions of SRT division algorithms are isomorphic to each other for all divisors.

DEFINITION 6 (expanding; see Bowen [2]). *We will say that a transformation T on an interval is expanding if it has the property that $\sup_{n>0} \mu(T^n U) = 1$ for all open intervals U with $\mu(U) > 0$, where μ is any normalized measure that is absolutely continuous with respect to Lebesgue measure.*

DEFINITION 7 (straddle). *Let U be an interval of reals (either open, closed, or half open) and let $p \in \mathbb{R}^+$. If $p \in U^\circ$,⁵ then we say that U straddles p .*

THEOREM 8. *The SRT division transformation is expanding for all real divisors.*

Proof. Let (X, \mathcal{B}, m) be a probability space, where $X = [0, 1)$, \mathcal{B} is the Borel σ -algebra on X , and m is the Lebesgue measure on \mathcal{B} .⁶ Let $T_D : X \rightarrow X$ be the SRT division transformation for a given normalized divisor D as defined in (1).

Let us define an infinite sequence of intervals $\mathcal{U} = \{U_i\}_{i \in \mathbb{N}}$ as

$$U_1 = U \quad \text{and}$$

$$U_{i+1} = \begin{cases} T_D(U_i) & : U_i^\circ \subseteq [0, \frac{1}{2}) \quad \text{or } U_i^\circ \subseteq [\frac{1}{2}, 1), \\ T_D(U_i \cap [0, \frac{1}{2})) & : U_i^\circ \not\subseteq [0, \frac{1}{2}) \text{ and } U_i^\circ \not\subseteq [\frac{1}{2}, 1) \text{ and} \\ & m(U_i \cap [0, \frac{1}{2})) \geq m(U_i \cap [\frac{1}{2}, 1)), \\ T_D(U_i \cap [\frac{1}{2}, 1)) & : U_i^\circ \not\subseteq [0, \frac{1}{2}) \text{ and } U_i^\circ \not\subseteq [\frac{1}{2}, 1) \text{ and} \\ & m(U_i \cap [0, \frac{1}{2})) < m(U_i \cap [\frac{1}{2}, 1)). \end{cases}$$

PROPERTY 1. *For all U_i such that $\frac{1}{2} \notin U_i^\circ$ and $D \notin U_i^\circ$, $m(U_{i+1}) = 2m(U_i)$.*

Proof. If a U_i° is a subset of either $[0, \frac{1}{2})$, $[\frac{1}{2}, D)$, or $[D, 1)$, then we are in the first case of the \mathcal{U} definition and we apply T_D directly. Since each of the three cases of the T_D expands an interval by a factor of two, it is clear that $m(T_D(U_i)) = m(U_{i+1}) = 2m(U_i)$.

PROPERTY 2. *For all U_i where $D \notin U_i$, $m(U_{i+1}) \geq m(U_i)$.*

Proof. Assume that $D \notin U_i$. If $\frac{1}{2} \notin U_i$, then according to Property 1, U_{i+1} doubles. Otherwise, $\frac{1}{2} \in U_i$, and therefore, to find U_{i+1} , we must consider the second and third cases of the \mathcal{U} sequence. In the worst case, $m(U_i \cap [0, \frac{1}{2})) = m(U_i \cap [\frac{1}{2}, D))$, and regardless of which half we choose, $m(U_i \cap [0, \frac{1}{2})) = m(U_i \cap [\frac{1}{2}, D)) = \frac{1}{2}m(U_i)$. By applying T_D to this truncated interval, we double what we halved so that $m(U_i) = m(U_{i+1})$.

By way of contradiction, let us assume that there exists an initial U such that the sequence \mathcal{U} never expands to fill X . Such a sequence can never include the point D ; if it did, there would be a small interval about D that would be mapped to $[0, \varepsilon)^\circ$, and this interval would quickly expand to fill the whole interval. We can show that the following property will hold.

PROPERTY 3. *There exists N such that for all $i \geq N$,*

- (a) $m(U_i \cap [0, \frac{1}{2})), m(U_i \cap [\frac{1}{2}, 1)) > 0$ (in other words, all subsequent intervals must straddle $\frac{1}{2}$), and
- (b) $m(U_i \cap [0, \frac{1}{2})) < m(U_i \cap [\frac{1}{2}, 1))$ (in other words, all subsequent U_i must be such that the right half of U_i is not discarded by the definition of \mathcal{U}).

Proof of Property 3(a). Property 1 says that the only way not to double is to straddle $\frac{1}{2}$. Therefore, at a minimum, it must be the case that $\frac{1}{2}$ is eventually included every time or else the interval will double a sufficient number of times to include D , which would be a contradiction.

Proof of Property 3(b). If $m(U_i \cap [0, \frac{1}{2})) \geq m(U_i \cap [\frac{1}{2}, 1))$, then we have $U_i = (\frac{1}{2} - \varepsilon, \frac{1}{2} + \varepsilon')$, where $\varepsilon > \varepsilon'$. Now $U_{i+1} = T_D(U_i) = T_D(\frac{1}{2} - \varepsilon, \frac{1}{2}) = (1 - 2\varepsilon, 1)$. But,

⁵The symbol \circ as the exponent of an interval denotes an open version of the interval.

⁶For an interval $[a, b]$, the Lebesgue measure is defined as $m([a, b]) = b - a$.

since D is not in U_{i+1} , $\frac{1}{2}$ cannot be in U_{i+1} and Property 3(a) fails, resulting in a contradiction.

By Property 3(a), we will eventually be in a situation where $U_i = (\frac{1}{2} - \varepsilon', \frac{1}{2} + \varepsilon)$, $\varepsilon' < \varepsilon$, and Property 3(a) will hold for every subsequent interval. So then

$$U_{i+1} = T_D(\frac{1}{2} - \varepsilon', \frac{1}{2} + \varepsilon) = T_D[\frac{1}{2}, \frac{1}{2} + \varepsilon) = (2D - 1 - 2\varepsilon, 2D - 1]$$

by Property 3(b). But again by Property 3(a),

$$U_{i+2} = T_D(2D - 1 - 2\varepsilon, 2D - 1] = T_D[\frac{1}{2}, 2D - 1] = [2 - 2D, 2D - 1].$$

It is now clear that $\frac{1}{2}$ is at the midpoint of U_{i+2} and that we must now pick the left half of the interval, which contradicts Property 3(b). Therefore, D will eventually be included in an interval and the sequence will expand to fill all of X . \square

We can now prove that the SRT division process is weak-mixing, and therefore Bernoulli, by two theorems of Bowen [2].

THEOREM 9 (see Bowen [2]). *Let T be a piecewise C^2 map of $[0, 1]$, μ be a smooth T -invariant probability measure, and $\lambda = \inf_{0 \leq x \leq 1} |f'(x)| > 1$. If the dynamical system (T, μ) is weak-mixing, then the natural extension of (T, μ) is Bernoulli.*

We mention here that the natural extension of (T, μ) is the associated automorphic transformation that we alluded to at the beginning of this section. See Petersen [16, p. 13] for an exact definition.

THEOREM 10 (see Bowen [2]). *With T and μ as in Theorem 9, (T, μ) will be weak-mixing if T is expanding.*

THEOREM 11 (see Lasota and Yorke [9]). *Let (X, \mathcal{B}, m) be a probability space and let $T : X \rightarrow X$ be a piecewise C^2 function such that $\inf |T'| > 1$. If P is the Perron–Frobenius operator associated with T , then for any $f \in L^1$, the sequence $(\frac{1}{n} \sum_{k=0}^{n-1} P^k f)_{n=1}^\infty$ is convergent in norm to a function $f^* \in L^1$. The limit function f^* has the property that $Pf^* = f^*$ and, consequently, the measure $d\mu^* = f^* dm$ is invariant under T .*

Having established that T_D is expanding, we now use the above three theorems to prove the central result of this paper.

THEOREM 12. T_D is Bernoulli.

Proof. From the definition of T_D , we see that T_D is C^2 and that

$$\inf_{0 \leq x \leq 1} |T_D'(x)| = 2 > 1$$

since $|T_D'(x)| = 2$ for all x , for which the derivative is defined. Since

$$\inf_{0 \leq x \leq 1} |T_D'(x)| > 1,$$

by Theorem 11 there exists at least one μ such that μ is a smooth T_D -invariant probability measure. By Theorem 8, we see that Theorem 10 holds. Hence, (T_D, μ) is weak-mixing, and by Theorem 9 (T_D, μ) is Bernoulli. \square

That all T_D are Bernoulli is a very useful property because we can use entropy as a complete invariant to show isomorphism among the two-sided Bernoulli shifts associated with T_D that have the same entropy. This comes from the contribution of Ornstein to the Kolmogorov–Ornstein theorem.

THEOREM 13 (see Kolmogorov [6, 7] and Ornstein [14]). *Two Bernoulli shifts are isomorphic if and only if they have the same entropy.*

Before we calculate the entropy, we review the definition of entropy for a transformation along with some supporting definitions that follow the development presented by Walters [22, pp. 75–87].

DEFINITION 14 (partition). *A partition of (X, \mathcal{B}, m) is a disjoint collection of nonempty elements of \mathcal{B} whose union is X .*

DEFINITION 15 (join). *Let \mathcal{P} and \mathcal{Q} be finite partitions of (X, \mathcal{B}, m) . Then $\mathcal{P} \vee \mathcal{Q} = \{P \cap Q : P \in \mathcal{P} \text{ and } Q \in \mathcal{Q}\} \setminus \{\emptyset\}$ is called the join of \mathcal{P} and \mathcal{Q} . Note that $\mathcal{P} \vee \mathcal{Q}$ is also a finite partition of (X, \mathcal{B}, m) .*

DEFINITION 16 (entropy of a partition). *Let (X, \mathcal{B}, m) be a probability space and let $\mathcal{P} = \{P_1, \dots, P_k\}$ be a finite partition of (X, \mathcal{B}, m) . The entropy of the partition is defined as*

$$H(\mathcal{P}) = - \sum_{i=1}^k m(P_i) \log m(P_i).$$

DEFINITION 17 (entropy of a transformation with respect to a partition). *Suppose $T : X \rightarrow X$ is a measure-preserving transformation of the probability space (X, \mathcal{B}, m) . If \mathcal{P} is a finite partition of (X, \mathcal{B}, m) , then*

$$h(T, \mathcal{P}) = \lim_{n \rightarrow \infty} \frac{1}{n} H \left(\bigvee_{i=0}^{n-1} T^{-i} \mathcal{P} \right)$$

is called the entropy of T with respect to partition \mathcal{P} .

DEFINITION 18 (entropy of a transformation). *Let $T : X \rightarrow X$ be a measure-preserving transformation of the probability space (X, \mathcal{B}, m) and suppose $h(T) = \sup h(T, \mathcal{P})$, where the supremum is taken over all finite partitions \mathcal{P} of (X, \mathcal{B}, m) . Then $h(T)$ is called the entropy of T .*

In general it can be very difficult to calculate the entropy for a class of transformations directly from the definition of entropy. Even with the many standard formulas that have been derived for calculating entropy, a great number of systems found in practice are not covered. Simple SRT division is one such dynamical system for which it is not easy to calculate the entropy from results found in standard textbooks on ergodic theory. Fortunately, a result by Ledrappier does allow us to calculate the entropy for simple SRT division.

The following definitions and theorems involving C-maps and PC-maps are taken from a paper of Ledrappier [10] and have been streamlined for our argument.

DEFINITION 19 (C-map; see Ledrappier [10]). *A real function f defined on an interval $[a, b]$ is said to be a C-map if f is continuously differentiable and its derivative f' has the following properties:*

- (a) *f' satisfies a Hölder condition⁷ of order $\varepsilon > 0$.*
- (b) *There are only a finite number of points $x \in [a, b]$ where $f'(x) = 0$. We denote them by $a \leq a_1 < a_2 < \dots < a_n \leq b$ with $f'(a_i) = 0$ for $0 < i \leq n$.*
- (c) *There exist positive numbers k_i^- (k_i^+) such that*

$$\left| \log \frac{|f'(x)|}{|x - a|^{k_i^{-(+)}}} \right|$$

is bounded in a left (right) neighborhood of a_i .

⁷A function $f(x)$ defined on an interval $[a, b]$ satisfies a Hölder condition of order $\varepsilon \in \mathbb{R}^+$ if there exists $c \in \mathbb{R}^+$ such that for any two points $p_1, p_2 \in [a, b]$, $|f(p_1) - f(p_2)| \leq c|p_1 - p_2|^\varepsilon$.

DEFINITION 20 (PC-map; see Ledrappier [10]). A map $f : [0, 1] \rightarrow [0, 1]$ is called a PC-map if there exists a finite partition $0 < b_1 < b_2 \cdots < b_m < 1$ such that f is a C-map from $[b_j, b_{j+1}]$ into $[0, 1]$ for any j .

THEOREM 21 (see Ledrappier [10]). Let f be a PC-map. If μ is an a.c.i.m. (absolutely continuous invariant measure), then

$$(5) \quad h(f) = \int \log |f'| \, d\mu.$$

This formula was shown by Rohlin [18] to hold for a smaller class of transformations, which does not include the T_D associated with SRT division.

THEOREM 22. The entropy $h(T_D)$ of T_D for $D \in [\frac{1}{2}, 1)$ is equal to $\int \log |T_D'| \, d\mu = \log 2$.

Proof. By the definition of a PC-map, T_D is a PC-map if each of the three functions $T_D|_{[0, \frac{1}{2}]}$, $T_D|_{[\frac{1}{2}, D)}$, and $T_D|_{[D, 1]}$ is a C-map.

Trivially, each T_D restricted to any of the three domains $[0, \frac{1}{2})$, $[\frac{1}{2}, D)$, and $[D, 1)$ satisfies a Hölder condition of order $\varepsilon = 1$ because each piece of T_D is just a line of slope two. Thus condition (a) of Definition 19 is satisfied. Conditions (b) and (c) are satisfied because there are no points for which the derivative is equal to zero within a given line segment. Thus each of the three segments of T_D are C-maps and, by Definition 20, T_D is a PC-map.

Now, since each T_D is Bernoulli, there exists a unique a.c.i.m. (call it μ) for each T_D . By Theorem 21, we can use (5) to calculate the entropy:

$$h(T_D) = \int \log |T_D'| \, d\mu = \log 2 \int d\mu = \log 2. \quad \square$$

With the proof of Theorem 22 we have established isomorphism among the automorphic transformations (or natural extensions) associated with simple SRT division transformations by an application of the Kolmogorov–Ornstein theorem. The key to obtaining this result was being able to show that T_D has Bowen’s expanding property. In the following section, we extend these results to a more general type of SRT division.

4. Multithreshold SRT division. A simple optimization to the original SRT division algorithm, at least with the historical concern of avoiding additions and subtractions in mind, is the inclusion of additional divisors to increase the shift average. In this section, we prove that all such division algorithms with reasonable assumptions on the separation of the divisor multiples have the expanding property. It will be useful to precisely define a class of “multithreshold” SRT transformations.

DEFINITION 23. Let $\alpha \in \mathbb{R}^n$ be such that

- (a) $0 < \alpha_1 < \alpha_2 < \cdots < \alpha_n$, and
- (b) for all $x, D \in [\frac{1}{2}, 1)$, there exists $i \in \{1, \dots, n\}$ such that $|\alpha_i D - x| < \frac{1}{2}$.

We define \mathfrak{A}_n to be the set of all $\alpha \in \mathbb{R}^n$ satisfying conditions (a) and (b). Also, $\mathfrak{A} = \bigcup_{n \in \mathbb{N}} \mathfrak{A}_n$.

DEFINITION 24 (peaks and valleys). Given an $\alpha \in \mathfrak{A}_{n \geq 2}$, the point of intersection between two lines $f(x) = 2(x - \alpha_i D)$ and $g(x) = 2(\alpha_{i+1} D - x)$ will be called a peak and is denoted by $\psi_i = (\frac{1}{2}D(\alpha_{i+1} + \alpha_i), D(\alpha_{i+1} - \alpha_i))$. For convenience, we will refer to the abscissa as $\psi_i^x = \frac{1}{2}D(\alpha_{i+1} + \alpha_i)$ and to the ordinate as $\psi_i^y = D(\alpha_{i+1} - \alpha_i)$. The point of intersection of the two lines $f(x) = 2(\alpha_i D - x)$ and $g(x) = 2(x - \alpha_i D)$ is $(\alpha_i D, 0)$ and will be called a valley.

DEFINITION 25. For a $D \in [\frac{1}{2}, 1)$ and an $\alpha \in \mathfrak{A}$, define the transformation $T_{D,\alpha}(x) : [0, 1) \rightarrow [0, 1)$. For $\alpha \in \mathfrak{A}_1$, we get the familiar transformation

$$T_{D,\alpha}(x) = \begin{cases} 2x & : 0 \leq x < \frac{1}{2}, \\ |2(D-x)| & : \frac{1}{2} \leq x < 1. \end{cases}$$

For $\alpha \in \mathfrak{A}_2$,

$$T_{D,\alpha}(x) = \begin{cases} 2x & : 0 \leq x < \frac{1}{2}, \\ |2(\alpha_1 D - x)| & : \frac{1}{2} \leq x < \psi_1^x, \\ |2(\alpha_2 D - x)| & : \frac{1}{2} \leq x \text{ and } \psi_1^x \leq x < 1. \end{cases}$$

For $\alpha \in \mathfrak{A}_{n \geq 3}$,

$$T_{D,\alpha}(x) = \begin{cases} 2x & : 0 \leq x < \frac{1}{2}, \\ |2(\alpha_1 D - x)| & : \frac{1}{2} \leq x < \psi_1^x, \\ |2(\alpha_i D - x)| & : \frac{1}{2} \leq x \text{ and } \psi_i^x \leq x < \psi_{i+1}^x, \\ |2(\alpha_n D - x)| & : \frac{1}{2} \leq x \text{ and } \psi_{n-1}^x \leq x < 1. \end{cases}$$

DEFINITION 26. Define $\mathfrak{M}_n = \{T_{D,\alpha} : D \in (\frac{1}{2}, 1], \alpha \in \mathfrak{A}_n\}$ and define $\mathfrak{M} = \bigcup_{n \in \mathbb{N}} \mathfrak{M}_n$. We call \mathfrak{M}_n the set of all n -threshold SRT division transformations and call \mathfrak{M} the set of multithreshold SRT division transformations.

Table 2 shows an example of dividing 0.67 by 0.75 using multithreshold SRT division with $\alpha = (0.75, 1, 1.25)$. This is the same example calculation as in Table 1, but here the dividend has been computed to twice as many digits of precision with the same effective number of uses of the adders. We say “effective” because in multithreshold SRT division, there are several adders working in parallel. In a real implementation of multithreshold SRT division, the values for α must be carefully chosen so that not too much overhead is required to select a good partial remainder. There is also a tradeoff between the amount of overhead in choosing a good partial remainder and the precision to which a good partial remainder is selected.

TABLE 2
An example of multithreshold SRT division.

$p_0 = 0.67$	$= 0.67$	$q_0 = \alpha_2$	$Q_0 = 1$
$p_1 = 2(0.67 - \alpha_2 D)$	$= -0.16$	$q_3 = -\alpha_1$	$Q_3 = 0.90625$
$p_4 = 2(2^2(-0.16) + \alpha_1 D)$	$= -0.155$	$q_6 = -\alpha_1$	$Q_6 = 0.89453125$
$p_7 = 2(2^2(-0.155) + \alpha_1 D)$	$= -0.115$	$q_{10} = -\alpha_3$	$Q_{10} \doteq 0.8933105469$
$p_{11} = 2(2^3(-0.115) + \alpha_3 D)$	$= 0.035$	$q_{15} = \alpha_1$	$Q_{15} \doteq 0.8933334351$
$p_{16} = 2(2^4(0.035) - \alpha_1 D)$	$= -0.005$	$q_{23} = -\alpha_1$	$Q_{23} \doteq 0.8933333456$
$p_{24} = 2(2^7(0.005) + \alpha_1 D)$	$= -0.155$		

Condition (b) in Definition 23 guarantees that the division algorithm generates a new quotient bit at every step. Although the condition makes sense intuitively, it is not immediately obvious just by inspection if an α satisfies the condition. Lemma 28 below provides an easier way to check.

LEMMA 27. If $\alpha = (\alpha_1)$, then condition (b) of Definition 23 is satisfied if and only if $\alpha_1 = 1$.

Proof. If $\alpha_1 = 1$, then $\max_{D,x \in [1/2,1)} |\alpha_1 D - x| < \frac{1}{2}$. Now consider the cases when $\alpha_1 \neq 1$ and $\varepsilon \in \mathbb{R}^+$. If $\alpha_1 = 1 + \varepsilon$, then when $D = \frac{1}{1+\varepsilon}$ and $x = \frac{1}{2}$, $|\alpha_1 D - x| =$

$1 - \frac{1}{2} = \frac{1}{2} \not\leq \frac{1}{2}$. On the other hand, if $\alpha_1 = 1 - \varepsilon$, then when $D = \frac{1}{2}$ and $x = 1 - \frac{\varepsilon}{2}$, $|\alpha_1 D - x| = 1 - \frac{\varepsilon}{2} - (1 - \varepsilon)\frac{1}{2} = \frac{1}{2} \not\leq \frac{1}{2}$. \square

LEMMA 28. An $\alpha \in \mathfrak{A}_n$ that satisfies condition (a) of Definition 23 also satisfies condition (b) if and only if for some $i, j \in \{1, \dots, n\}$ (possibly $i = j$) either

- (i) $\alpha_i \in (0, \frac{1}{2}]$ and $\alpha_j \in [1, 1 + \alpha_i]$, or
- (ii) $\alpha_i \in [\frac{1}{2}, 1]$ and $\alpha_j \in [1, 3\alpha_i]$.

Proof (sketch). Lemma 27 has shown that a single component α of α with $\alpha = 1$ is sufficient to ensure that the range of $f(x) = 2|\alpha D - x|$ is equal to $[0, 1)$ as x and D range over $[\frac{1}{2}, 1)$. It is easy to see, based on the proof of Lemma 27, that if there does not exist $i \in \{1, \dots, n\}$ such that $\alpha_i = 1$, then there must exist $i, j \in \{1, \dots, n\}$ ($i < j$), where $\alpha_i < 1$ and $\alpha_j > 1$.

Let us assume that i is the largest value, where $\alpha_i < 1$, and let us assume that j is the smallest value, where $\alpha_j > 1$ (therefore $j = i + 1$). We make these assumptions because no other scalars of D will have an influence on whether or not condition (b) is satisfied. Consider the case when $\alpha_i \in (0, \frac{1}{2}]$. In this case, when D is close enough to 1, some of the line $f(x) = 2(x - \alpha_i D)$ appears in the region (denoted R), when $\frac{1}{2} \leq x < 1, 0 \leq T_\alpha(x) < 1$. When a portion of the line $f(x)$ appears in region R , we must put restrictions on α_j in terms of α_i so that the peak ψ_1 is always in R . ψ_i^y is greatest when $D = 1$. We find the maximum allowable value of α_j by setting $D = 1$ and solving $\psi_i^y = 1$ for α_j :

$$\psi_i^y = 1 \Rightarrow D(\alpha_j - \alpha_i) = 1 \Rightarrow \alpha_j = \alpha_i + 1.$$

Therefore, if $\alpha_i \in (0, \frac{1}{2}]$, then $\alpha_j \in [1, 1 + \alpha_i]$.

In the case when $\alpha_i \in [\frac{1}{2}, 1]$, for large enough values of D , the line $f(x) = 2(x - D\alpha_i)$ crosses the line $x = 1$ in the range $[0, 1)$. Because of this, we must loosen the restriction that $\alpha_j \in [1, 1 + \alpha_j]$. It is straightforward to calculate that $f(x)$ begins to cross the line $x = 1$ in the range $[0, 1)$ when $D = \frac{1}{2\alpha_i}$. By solving $\psi_i^y = 1$ for α_j when $D = \frac{1}{2\alpha_i}$ we can ensure that as D becomes smaller, the peak ψ_i will always be in region R :

$$\psi_i^y = 1 \Rightarrow D(\alpha_j - \alpha_i) = 1 \Rightarrow \frac{1}{2\alpha_i}(\alpha_j - \alpha_i) = 1 \Rightarrow \alpha_j = 3\alpha_i.$$

Therefore, if $\alpha_i \in [\frac{1}{2}, 1]$, then $\alpha_j \in [1, 3\alpha_i]$. \square

DEFINITION 29 (separation). For $\alpha \in \mathfrak{A}_n$, we define the separation in α as

$$sep(\alpha) = \max_{i \in \{1, \dots, n-1\}} \frac{\alpha_{i+1}}{\alpha_i}.$$

Limiting the separation is a convenient way to restrict the subset of \mathfrak{A} being considered. If $sep(\alpha) = r$, we say that “the divisor multiples in α are separated by at most a factor of r .”

We are now ready to show that all multithreshold SRT division transformations are Bernoulli, given a necessary restriction on the multiples of the divisor. As in the case for a single divisor, it will be useful to define a sequence of intervals that are subsets of the sequence of sets that would arise from repeatedly applying $T_{D,\alpha}$ to an initial open interval. Unless otherwise noted, assume that the function m denotes the Lebesgue measure.

DEFINITION 30. Given an initial open interval $U \subset [0, 1)$ and $T_{D,\alpha} \in \mathfrak{M}$, we

define the infinite sequence of intervals $\mathcal{U} = \{U_i\}_{i \in \mathbb{N}}$ as

$$U_1 = U \quad \text{and}$$

$$U_{i+1} = \begin{cases} T_{D,\alpha}(U_i) & : \quad U_i^\circ \subseteq [0, \frac{1}{2}) \quad \text{or } U_i^\circ \subseteq [\frac{1}{2}, 1), \\ T_{D,\alpha}(U_i \cap [0, \frac{1}{2})) & : \quad U_i^\circ \not\subseteq [0, \frac{1}{2}) \text{ and } U_i^\circ \not\subseteq [\frac{1}{2}, 1) \text{ and} \\ & \quad m(U_i \cap [0, \frac{1}{2})) \geq m(U_i \cap [\frac{1}{2}, 1)), \\ T_{D,\alpha}(U_i \cap [\frac{1}{2}, 1)) & : \quad U_i^\circ \not\subseteq [0, \frac{1}{2}) \text{ and } U_i^\circ \not\subseteq [\frac{1}{2}, 1) \text{ and} \\ & \quad m(U_i \cap [0, \frac{1}{2})) < m(U_i \cap [\frac{1}{2}, 1)). \end{cases}$$

DEFINITION 31 (critical points). For a given $T_{D,\alpha}$ where $\alpha \in \mathfrak{A}_n$, define the set

$$C = \{c_i : i \in \{1, \dots, m\}, c_i \in B \cup \{0, \frac{1}{2}, 1\}\}$$

where $B = \{b : \frac{1}{2} < b < 1 \text{ and } b \in \{\alpha_1 D, \dots, \alpha_n D\} \cup \{\psi_1^x, \dots, \psi_{n-1}^x\}\}$ and $c_1 < c_2 < \dots < 1$. C is called the set of critical points for $T_{D,\alpha}$.

LEMMA 32 (doubling). Given $T_{D,\alpha} \in \mathfrak{M}$, let the sequence of intervals \mathcal{U} be defined as in Definition 30 and let U_i be some interval in the sequence. Furthermore, let $C = \{c_1, \dots, c_m\}$ be the set of critical points for $T_{D,\alpha}$. If $U_i \subseteq [c_j, c_{j+1}]$ for some $j \in \{1, \dots, m-1\}$, then $m(U_{i+1}) = 2m(U_i)$.

Proof. Since $U_i \subseteq [c_j, c_{j+1}]$ for some $j \in \{1, \dots, m-1\}$, because we are in the first case of the definition of \mathcal{U} , either $U_i^\circ \subseteq [0, \frac{1}{2})$ or $U_i^\circ \subseteq [\frac{1}{2}, 1)$. By simple inspection of the individual cases that define $T_{D,\alpha}$, it is apparent that all of U_i , except possibly the points c_j and c_{j+1} , fall within the same case of $T_{D,\alpha}$. Therefore, the resulting interval U_{i+1} will be double the length of U_i . \square

DEFINITION 33 (active valleys). Given $T_{D,\alpha} \in \mathfrak{M}_n$, define

$$V = \{\alpha_i D : i \in \{1, \dots, n\} \text{ and } \frac{1}{2} < \alpha_i D < 1\}.$$

V is called the set of active valleys for $T_{D,\alpha}$.

DEFINITION 34 (active peaks). Given $T_{D,\alpha} \in \mathfrak{M}_n$, define

$$P = \{\psi_i^x : i \in \{1, \dots, n-1\} \text{ and } \frac{1}{2} < \psi_i^x < 1\}.$$

P is called the set of active peaks for $T_{D,\alpha}$.

LEMMA 35 (nonshrinking). Given $T_{D,\alpha} \in \mathfrak{M}_n$ with $\text{sep}(\alpha) \leq \frac{5}{3}$, let the sequence of intervals $\{U_i\}_{i \in \mathbb{N}}$ be defined as above and let V denote the set of active valleys for $T_{D,\alpha}$. For any interval $U_i \in \mathcal{U}$ such that $V \cap U_i = \emptyset$, either $m(U_{i+1}) \geq m(U_i)$ or $m(U_{i+2}) \geq m(U_i)$.

Proof. $\text{sep}(\alpha) \leq \frac{5}{3}$ implies that $\alpha_{i+1} \leq \frac{5}{3}\alpha_i$. For a given separation, the value of ψ_i^y is maximized when $\psi_i^x = 1$. This implies that $\alpha_i = \frac{3}{4D}$. We calculate the value of ψ_i^y with the assumption that $\psi_i^x = 1$ to get a bound on ψ_i^y for $D < 1$:

$$\psi_i^y \leq D(\frac{5}{3}\alpha_i - \alpha_i) = D(\frac{2}{3}\alpha_i) = D(\frac{2}{3} \frac{3}{4D}) = \frac{1}{2}.$$

Case 1. Consider when $U_i \subseteq [0, \frac{1}{2}]$. In this case, $m(U_{i+1}) = 2m(U_i)$.

Case 2. Consider when $U_i \subseteq [\frac{1}{2}, 1)$. The interval U_i can span at most one peak. Therefore, $m(U_{i+1}) \geq m(U_i)$. A further observation is that since $U_{i+1} \subseteq [0, \frac{1}{2}]$, $m(U_{i+2}) = 2m(U_{i+1})$.

Case 3. Consider when $U_i \not\subseteq [0, \frac{1}{2}]$ and $U_i \not\subseteq [\frac{1}{2}, 1)$. In this case, U_i straddles $\frac{1}{2}$. From the definition of \mathcal{U} , we see that in the worst case we might throw away up to half of U_i . Call the part not thrown away U_i' and observe that $m(U_i') \geq \frac{1}{2}m(U_i)$. Now, either $U_i' \subseteq [0, \frac{1}{2}]$ or $U_i' \subseteq [\frac{1}{2}, 1)$. If $U_i' \subseteq [0, \frac{1}{2}]$, then $m(U_{i+1}) = 2m(U_i') \geq m(U_i)$. If $U_i' \subseteq [\frac{1}{2}, 1)$, then $m(U_{i+2}) = 2m(U_i') \geq m(U_i)$. \square

LEMMA 36. *A multithreshold SRT division transformation $T_{D,\alpha} \in \mathfrak{M}$ is expanding when $\text{sep}(\alpha) \leq \frac{5}{3}$.*

Proof. Let V be the set of active valleys (as defined in Definition 33) for a $T_{D,\alpha}$. Let P be the set of active peaks (as defined in Definition 34) for a $T_{D,\alpha}$. Let $\mathcal{U} = \{U_i\}_{i \in \mathbb{N}}$ be the sequence of intervals associated with a $T_{D,\alpha}$ and an initial interval U .

By way of contradiction, assume that a $T_{D,\alpha}$ is not expanding. This means that for some $T_{D,\alpha}$, there does not exist an interval U_i where any of the points in V are contained in U_i . This is true because if any of the valley points are in U_i , then $U_{i+1} = [0, \varepsilon)$ or $U_{i+1} = [0, \varepsilon]$, and after a finite number of steps, U_i will have grown to include all of $[0, 1)$.

If there is a sequence \mathcal{U} that avoids all points in V , then by Lemma 35 it must be true that the intervals in the sequence can double only a finite number of times. Let $i \in \mathbb{N}$ be the first index for which there is no $j > i$ such that $m(U_j) \geq 2m(U_i)$. It now follows that U_i straddles $\frac{1}{2}$. The proof for Lemma 35 reveals that this is the only situation where it is not necessarily the case that either $m(U_{i+1}) = 2m(U_i)$ or $m(U_{i+2}) = 2m(U_i)$. In fact, U_i must straddle both $\frac{1}{2}$ and $\min P$. If $\min P$ is not straddled and $m(U_i \cap [0, \frac{1}{2})) < m(U_i \cap [\frac{1}{2}, 1))$, then either $m(U_{i+2}) \geq 2m(U_i)$ or $m(U_{i+3}) \geq 2m(U_i)$. In the other possibility, where $\min P$ is not straddled and $m(U_i \cap [0, \frac{1}{2})) \geq m(U_i \cap [\frac{1}{2}, 1))$, we find that $m(U_{i+2}) \geq 2m(U_i)$.

Assuming that U_i straddles both $\frac{1}{2}$ and $\min P$, we also observe that there can be no $j > i$ such that $m(U_j \cap [0, \frac{1}{2})) \geq m(U_j \cap [\frac{1}{2}, 1))$ because this quickly leads to doubling. In other words, the right side must be larger than the left side whenever we straddle $\frac{1}{2}$. Therefore, we must be in the situation where

$$\begin{aligned} U_i &= (\frac{1}{2} - \varepsilon', \frac{1}{2} + \varepsilon), \quad \varepsilon' < \varepsilon \\ \Rightarrow U_{i+1} &= (\min\{2(\frac{1}{2} - \alpha_i D), 2(\alpha_{i+1} D - (\frac{1}{2} + \varepsilon))\}, \psi_i^y) \\ \Rightarrow U_{i+2} &= (2 \min\{2(\frac{1}{2} - \alpha_i D), 2(\alpha_{i+1} D - (\frac{1}{2} + \varepsilon))\}, 2\psi_i^y) \\ \Rightarrow U_{i+3} &= (\min\{2(\frac{1}{2} - \alpha_i D), 2(\alpha_{i+1} D - 2\psi_i^y)\}, \psi_i^y) \\ \Rightarrow U_{i+4} &= (2 \min\{2(\frac{1}{2} - \alpha_i D), 2(\alpha_{i+1} D - 2\psi_i^y)\}, 2\psi_i^y) \\ \Rightarrow U_{i+5} &= (\min\{2(\frac{1}{2} - \alpha_i D), 2(\alpha_{i+1} D - 2\psi_i^y)\}, \psi_i^y) = U_{i+3}. \end{aligned}$$

It is apparent that the interval represented by U_{i+4} will re-occur every other interval ad infinitum. We now use this interval to show that in fact such a sequence of nonexpanding intervals is not possible.

Since U_{i+4} straddles $\frac{1}{2}$, we can compare the length of the left and right sides of U_{i+4} . Let $R = [\frac{1}{2}, 2\psi_i^y)$ denote the right side and let $L = (4(\frac{1}{2} - \alpha_i D), \frac{1}{2})$ and $L' = (4(\alpha_{i+1} D - 2\psi_i^y), \frac{1}{2})$ denote the two possibilities for the left side. The length of the right side is

$$m(R) = 2\psi_i^y - \frac{1}{2},$$

while the length of the left side is the larger of two possible lengths

$$m(L) = \frac{1}{2} - 4(\frac{1}{2} - \alpha_i D)$$

and

$$m(L') = \frac{1}{2} - 4(\alpha_{i+1}D - 2\psi_i^y) .$$

We then compare the differences between the right side and each of the two possible left sides. The first possibility is

$$\begin{aligned} m(R) - m(L) &= 2\psi_i^y - \frac{1}{2} - (\frac{1}{2} - 4(\frac{1}{2} - \alpha_i D)) \\ &= 2D(\alpha_{i+1} - \alpha_i) - 1 + 2 - 4\alpha_i D \\ &= 2\alpha_{i+1}D - 6\alpha_i D + 1 , \end{aligned}$$

while the second possibility is

$$\begin{aligned} m(R) - m(L') &= 2\psi_i^y - \frac{1}{2} - (\frac{1}{2} - 4(\alpha_{i+1} - 2\psi_i^y)) \\ &= 2D(\alpha_{i+1} - \alpha_i) - 1 + 4(\alpha_{i+1}D - 2D(\alpha_{i+1} - \alpha_i)) \\ &= -2\alpha_{i+1}D + 6\alpha_i D - 1 . \end{aligned}$$

It is now clear that

$$m(R) - m(L) = -(m(R) - m(L')) .$$

But this means that the length of the left side is always greater than or equal to the length of the right side, which contradicts our assumption that the right side must be bigger than the left side whenever the interval straddles $\frac{1}{2}$. \square

THEOREM 37. $T_{D,\alpha} \in \mathfrak{M}$ is Bernoulli when $sep(\alpha) \leq \frac{5}{3}$.

Proof. Let $T = T_{D,\alpha}$. From the definition of T , we see that $T_{D,\alpha}$ is C^2 and that $\inf_{0 \leq x \leq 1} |T'(x)| = 2 > 1$ since $|T'(x)| = 2$ for all x for which the derivative is defined. Since $\inf_{0 \leq x \leq 1} |T'(x)| > 1$, by Theorem 11, there exists at least one μ such that μ is a smooth T -invariant probability measure. By Lemma 36 we see that Theorem 10 holds when $sep(\alpha) \leq \frac{5}{3}$. Hence, (T, μ) is weak-mixing and, by Theorem 9, (T, μ) is Bernoulli when $sep(\alpha) \leq \frac{5}{3}$. \square

The calculation for entropy in multithreshold SRT division follows the same method used for single divisor SRT division. We begin by showing that $T_{D,\alpha}$ is a PC-map.

LEMMA 38. $T_{D,\alpha} \in \mathfrak{M}$ is a PC-map (as defined in Definition 20).

Proof. By inspection, each $T_{D,\alpha}$ is a finite collection of line segments, each with slope 2. Each of these line segments is a C-map by the same argument used in the proof for Theorem 22. Therefore, by definition, each $T_{D,\alpha}$ is a PC-map. \square

THEOREM 39. The entropy of any $T_{D,\alpha} \in \mathfrak{M}$ with $sep(\alpha) \leq \frac{5}{3}$ is $\log 2$.

Proof. By Lemma 38, all $T_{D,\alpha} \in \mathfrak{M}$ are PC-maps. By Theorem 37, $T_{D,\alpha}$ is Bernoulli when $sep(\alpha) \leq \frac{5}{3}$, and hence there exists a unique a.c.i.m. μ . Theorem 21 says that Rohlin’s formula for the entropy is true, and therefore

$$h(T_{D,\alpha}) = \int \log |T_{D,\alpha}'| d\mu = \log 2 \int d\mu = \log 2. \quad \square$$

5. Some restrictions on α for multithreshold division. In section 4, we showed that for all $T_{D,\alpha} \in \mathfrak{M}$, if $sep(\alpha) \leq \frac{5}{3}$, then $T_{D,\alpha}$ is Bernoulli. In this section, we construct examples of $T \in \mathfrak{M}_n$, for every n , that fail to be Bernoulli when the restriction that $sep(\alpha) \leq \frac{5}{3}$ is relaxed.

THEOREM 40. For $T_{D,\alpha} \in \mathfrak{M}_{n \geq 4}$, if $sep(\alpha) > \frac{5}{3}$, then for each $D \in [\frac{1}{2}, 1)$, there exist uncountably many α for which $T_{D,\alpha}$ is not ergodic.

Proof. We begin this proof by considering $T \in \mathfrak{M}_{n=4}$.

Assume that we relax the restrictions on α by $\varepsilon > 0$. This means that $sep(\alpha) \leq \frac{5}{3} + \varepsilon$ and that no peak can be above the line $f(x) = \frac{4+6\varepsilon x}{8+3\varepsilon}$. With this relaxation, we can define $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ with respect to a given D so that a subset of $[0, 1]$ is nonexpanding. We let $\alpha_1 = \frac{30+27\varepsilon}{80D+48D\varepsilon}$, $\alpha_2 = \frac{50+57\varepsilon}{80D+48D\varepsilon}$, $\alpha_3 = \frac{30-9\varepsilon}{40D+24D\varepsilon}$, and $\alpha_4 = \frac{50+21\varepsilon}{40D+24D\varepsilon}$. For our constructed α to be valid, we must be careful that conditions (a) and (b) of Definition 23 hold. Condition (a) requires that the components of α remain in ascending order. This is satisfied when $\varepsilon \in (0, \frac{2}{15}]$. Since ordering is maintained, $sep(\alpha) < 3$, and $\min_{D \in [1/2, 1], \varepsilon \in (0, 2/15]} \alpha_4 = 1.2 \geq 1$, to verify that condition (b) of Definition 23 holds, it is sufficient to show (by Lemma 28) that for all values of D and ε , either α_1, α_2 , or $\alpha_3 \in [\frac{1}{2}, 1]$. By maximizing and minimizing over ε and D , we find that $\alpha_1 \in [0.375, 0.7]$ and $\alpha_2 \in [0.625, 1.3]$. Figure 6 provides a visual proof that as ε is varied over $[0, \frac{2}{15}]$ and D is varied over $[\frac{1}{2}, 1]$, it is never the case that both $\alpha_1 \leq \frac{1}{2}$ and $\alpha_2 \geq 1$. Therefore, it is always the case that either α_1 or $\alpha_2 \in [\frac{1}{2}, 1]$.

Having verified that our defined α satisfies Definition 23, we calculate that peak $\psi_1 = (\frac{20+21\varepsilon}{40+24\varepsilon}, \frac{10+15\varepsilon}{40+24\varepsilon})$ and peak $\psi_3 = (\frac{20+3\varepsilon}{20+12\varepsilon}, \frac{10+15\varepsilon}{20+12\varepsilon})$. With this definition for α , and our assumption that $\varepsilon \in [0, \frac{2}{15}]$, the point ψ_3 will always touch the line $f(x)$ while remaining above the line $g(x) = \frac{1}{2}$, and the point ψ_1 will always be slightly below $f(x)$ while remaining above the line $g(x) = \frac{1}{4}$. All of the definitions have been chosen so that $1 - \psi_3^x = \psi_3^y - \frac{1}{2} = 2(\psi_1^x - \frac{1}{2}) = 2(\psi_1^y - \frac{1}{4})$. Another important feature in this construction is the interval between $\alpha_2 D$ and $\alpha_3 D$. Since ψ_2 is not used in our construction, it is possible to insert an arbitrary number of divisor multiples between $\alpha_2 D$ and $\alpha_3 D$. Thus, the results in this proof apply to $T \in \mathfrak{M}_n$ for arbitrarily large n .

We are now in a position to show that there exists a set of points A with $0 < m(A) < 1$, for which $T_{D,\alpha}(A) = A$. This is the definition of a transformation being nonergodic [8, p. 59]. Define $A = A_1 \cup A_2 \cup A_3$, where $A_1 = [\frac{1}{4} - (\psi_1^x - \frac{1}{2}), \frac{1}{4} + (\psi_1^x - \frac{1}{2})]$, $A_2 = [\frac{1}{2} - 2(\psi_1^x - \frac{1}{2}), \frac{1}{2} + 2(\psi_1^x - \frac{1}{2})]$, and $A_3 = [1 - 2(1 - \psi_3^x), 1]$. It can be shown by calculation that $T_{D,\alpha}(A_1) = A_2$, $T_{D,\alpha}(A_2) = A_1 \cup A_3$, and $T_{D,\alpha}(A_3) = A_2$. Therefore, $T_{D,\alpha}(A) = A$, and by definition, $T_{D,\alpha}$ is nonergodic or nonexpanding. \square

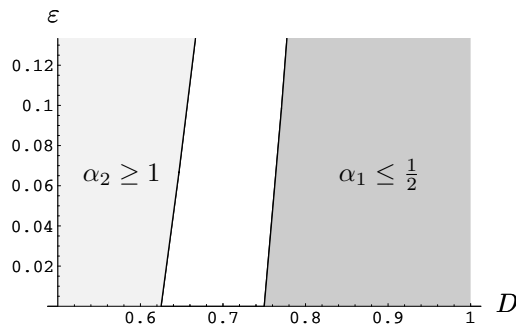


FIG. 6. Combined plot of the regions where $\alpha_1 \leq \frac{1}{2}$ and $\alpha_2 \geq 1$.

Figure 7 illustrates the type of transformation that we have constructed in the proof of Theorem 40. In this figure, $n = 4$, $D = \frac{11}{16}$, $\alpha = (\frac{37}{66}, \frac{21}{22}, 1, \frac{59}{33})$, and $sep(\alpha) = \frac{5}{3} + \frac{5}{51}$. The thick lines represent $T_{D,\alpha}$. The coarse dashed line represents the necessary separation restriction on α to guarantee that $T_{D,\alpha}$ is ergodic. In this case, partial remainders in the set $A = [\frac{11}{48}, \frac{13}{48}] \cup [\frac{22}{48}, \frac{26}{48}] \cup [\frac{44}{48}, 1)$ are mapped back to A by $T_{D,\alpha}$.

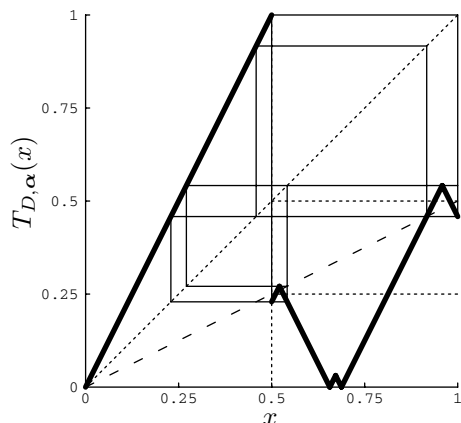


FIG. 7. An example of a nonergodic system for $T_{D,\alpha} \in \mathfrak{M}_{n \geq 4}$.

THEOREM 41. For $T_{D,\alpha} \in \mathfrak{M}_3$, if $sep(\alpha) \geq \frac{9}{5}$, then for each $D \in [\frac{1}{2}, 1)$, there exists an α for which $T_{D,\alpha}$ is not ergodic.

Proof. The proof for this theorem comes as a special case from the proof for Theorem 40. Consider $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ as defined in the proof for Theorem 40. When $sep(\alpha) = \frac{9}{5} = \frac{5}{3} + \frac{2}{15}$, we are in the special situation where $\alpha_2 = \alpha_3$. Since all of the results for the proof of Theorem 40 still hold, we now have an example transformation T with only three unique multiples of D , and this T has been proven to be nonergodic. \square

Figure 8 gives an example of a nonergodic transformation for $D = \frac{7}{12}$ and $\alpha = (\frac{2}{3}, \frac{8}{7}, \frac{44}{21})$. In this case, partial remainders in the set $A = [\frac{4}{18}, \frac{5}{18}] \cup [\frac{8}{18}, \frac{10}{18}] \cup [\frac{16}{18}, 1)$ are mapped back to A by $T_{D,\alpha}$.

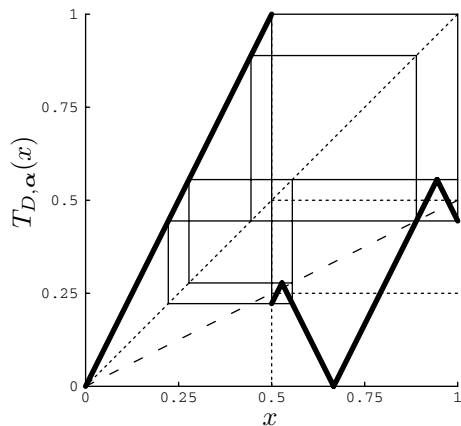


FIG. 8. An example of a nonergodic system for $T_{D,\alpha} \in \mathfrak{M}_3$.

THEOREM 42. For $T_{D,\alpha} \in \mathfrak{M}_2$, if $sep(\alpha) > 3$, then for some $D \in (\frac{1}{2}, 1)$, there exist uncountably many α for which $T_{D,\alpha}$ is not ergodic.

Proof. Assume that $sep(\alpha) \leq 3 + \varepsilon$ and $D \in (\frac{1}{2}, \frac{2+\varepsilon}{4})$. First, we choose $\alpha_1 = \frac{1}{4D}$ so that $\alpha_1 D = \frac{1}{4}$ and $\alpha_2 = 1 + \alpha_1$. Our restriction on D in terms of ε has been chosen so that $\alpha_2/\alpha_1 < 3 + \varepsilon$ when $\alpha_2 = 1 + \alpha_1$. Since $\alpha_2 > \alpha_1$, condition (a) of Definition

23 is satisfied. Since $\alpha_1 \in (\frac{1}{4}, \frac{1}{2})$ and $\alpha_2 \in (1, 1 + \alpha_1]$ by Lemma 28, condition (b) of Definition 23 is satisfied. Thus, our defined α is always valid. Define $A = [\frac{1}{2}, D]$. We now apply $T = T_{D,\alpha}$ to A :

$$\begin{aligned} T[\frac{1}{2}, D] &= [\min\{2(\frac{1}{2} - \alpha_1 D), 2(\alpha_2 D - D)\}, \psi_1^y] \\ &= [\min\{2(\frac{1}{2} - \frac{D}{4D}), 2(D + \frac{1}{4} - D)\}, D(\alpha_2 - \alpha_1)] \\ &= [\min\{\frac{1}{2}, \frac{1}{2}\}, D(1 + \frac{1}{4D} - \frac{1}{4D})] \\ &= [\frac{1}{2}, D]. \end{aligned}$$

Now, since $\frac{1}{2} < D < 1$, $0 < m(A) < 1$, and $T_{D,\alpha}A = A$, by definition $T_{D,\alpha}$ is not ergodic. \square

Figure 9 shows an example of a nonergodic system for $T_{D,\alpha} \in \mathfrak{M}_2$. In this example, $D = \frac{3}{5}$ and $\alpha = (\frac{5}{12}, \frac{17}{12})$. Partial remainders within the interval $[\frac{1}{2}, \frac{3}{5}]$ map back to $[\frac{1}{2}, \frac{3}{5}]$.

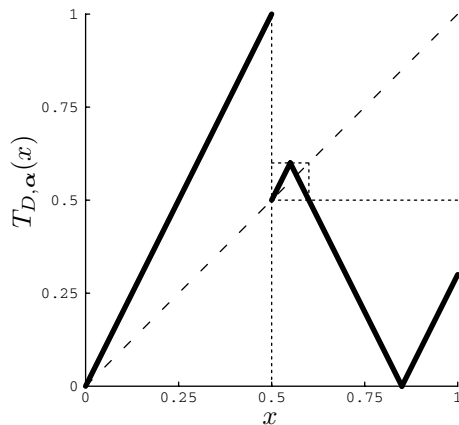


FIG. 9. An example of a nonergodic system for $T_{D,\alpha} \in \mathfrak{M}_2$.

6. Conclusions. The original question that inspired this paper was, “Is simple SRT division ergodic for all real divisors?” In pursuing the solution to this problem, we discovered that not only is simple SRT division ergodic for all divisors, but it is also Bernoulli. Having established a Bernoulli property, and having calculated the entropy for our transformations, we were able to use the Kolmogorov–Ornstein theorem to conclude that our transformations are equivalent to each other in the sense that their natural extensions are isomorphic. In proving these important properties for simple SRT division, we made extensive use of more general results from dynamical systems theory. Consequently, our results were shown to be easily extensible to more general division systems. In general, it is difficult to prove that a particular class of transformations is ergodic or Bernoulli. Our results have provided an effective means of proving both of these properties for a large class of SRT-like division algorithms.

From the standpoint of understanding an algorithm’s expected performance, it is necessary to know that when a stationary distribution is found, it is unique. Having established the uniqueness of stationary distributions, the next step is to find the actual stationary distribution for as wide a class of transformations as possible. In section 2, we gave the stationary distribution function for T_D , where $D \in [\frac{3}{4}, 1)$. Many

of the stationary distribution functions have been classified by Shively and Freiman for $D \in [\frac{3}{5}, \frac{3}{4}]$, although the derivations are not nearly as simple as for $D \in [\frac{3}{4}, 1)$. It turns out that things become very complicated when $D \in [\frac{1}{2}, \frac{3}{5}]$. In his thesis [20], Shively shows many interesting properties for the stationary distribution functions in this region. For example, he shows that there are many different intervals of D , where there are an infinite number of different stationary distribution equations. As such, the graph of the shift average for $D \in [\frac{1}{2}, \frac{3}{5}]$ is far from complete and appears to have a complex pattern (from the few points that have been plotted in this region). This is surprising, considering the simplicity of the underlying transformation. A better understanding of this final region of simple SRT division would be an interesting goal to pursue.

In the work of Freiman [5], it was first shown that the shift average for $D \in [\frac{3}{5}, \frac{3}{4}]$ is constantly 3, which can be easily shown to be the maximum possible shift average. This property was then used by Metze [12] to obtain a version of SRT division that has an expected shift average of 3 for all divisors. Another area to pursue would be to explore shift averages for multithreshold SRT division and, if other plateaus are found, they could possibly be used to obtain higher expected shift averages for all possible divisors. Undoubtedly, obtaining a complete understanding of the stationary distribution functions for multithreshold division would be even more difficult than it is for simple SRT division. It is possible that such results in this area could lead to improvements in modern SRT division. Related to this, it would be interesting to attempt to extend the results of this paper to modern SRT division.

Appendix. SRT division is chaotic. In section 2 we mentioned that SRT division is chaotic, and we prove this fact here for SRT division with the expansion property (see Definition 6). Simple SRT division was shown to be expanding for all divisors (see Theorem 8). Multithreshold SRT division, of which simple SRT division is a special case, was shown to be expanding when there is a restriction on the separation of the divisor multiples (see Lemma 36).

Although there are several definitions for what it means to be “chaotic,” the one given by Devaney [4] is commonly used. The following definitions are taken from [4, pp. 49, 50].

DEFINITION 43 (sensitivity). *A transformation $f : X \rightarrow X$ has sensitive dependence on initial conditions if there exists $\delta > 0$ such that, for any $x \in X$ and any neighborhood N about x , there exists $y \in X$ and $n \geq 0$ such that $|f^n(x) - f^n(y)| > \delta$.*

DEFINITION 44 (topological transitivity). *A transformation $f : X \rightarrow X$ is topologically transitive on X if, for any pair of open sets $U, V \subset X$, there exists n such that $f^n(U) \cap V \neq \emptyset$.*

DEFINITION 45 (chaotic). *Let $f : X \rightarrow X$ be a transformation of the set X . f is said to be chaotic on X if*

- (a) *f has sensitive dependence on initial conditions;*
- (b) *f is topologically transitive;*
- (c) *periodic points are dense in X .*

THEOREM 46. *Every multithreshold SRT division transformation with the expansion property is chaotic.*

Proof. Let T be a multithreshold SRT division transformation on $[0, 1)$ having the expansion property.

To prove sensitivity to initial conditions, we notice, by the expansion property, that for any open neighborhood N about a point x , the successive images of N under T expand to be arbitrarily close to filling the entire interval densely after a finite

number n of steps. In particular, since $f^n(x)$ must be at least distance $\frac{1}{2}$ away from either 0 or 1, for any $\frac{1}{2} > \delta > 0$, there is $y \in N$ such $|f^n(x) - f^n(y)| > \delta$. Therefore, T is sensitive to initial conditions on $[0, 1)$.

Again, by the expansion property of T , the successive images of any two open intervals will intersect within a finite number of steps. Therefore, T is topologically transitive on $[0, 1)$.

The expansion property alone is not sufficient to ensure that periodic points are dense. However, the existence of a particular sequence \mathcal{U} of nonempty subintervals as given in Definition 30 is sufficient, in general, for dense periodic points.

Let V_1 be an open interval on part of the domain where T is continuous. In the proof of Lemma 36, we showed that sequence $\{T^i(V_1)\}_{i=1}^\infty$ necessarily expands after a finite number of steps to fill the unit interval by proving that another sequence $\mathcal{U} = \{U_i\}$ of intervals expands to fill the unit interval where $U_i \subseteq V_i$ for $i \geq 0$. For a given U_i , U_{i+1} is chosen to be the largest subinterval of $T(U_i)$ on a continuous part of T 's domain. In the event of equal-length subintervals, we choose the left half to be U_i 's successor. \mathcal{U} induces another (possibly not unique) sequence of intervals $\{I_i\}$, where I_i is any subinterval of U_i , where $T(I_i) = U_{i+1}$. In any situation where part of $T(U_i)$ is discarded, there exists at least one interval $I_i \subset U_i$, where $T(I_i) = U_{i+1}$ because U_i is chosen such that T is continuous on U_i . Note that $T|_{I_i} : I_i \rightarrow U_{i+1}$ is a continuous onto map. Eventually, by the expansion property, it must be true that $U_n \supseteq U_1 = V_1$ for some n . We now have a sequence of continuous onto maps $T|_{U_1} : U_1 \rightarrow U_2$, $T|_{U_2} : U_2 \rightarrow U_3$, \dots , $T|_{U_{n-1}} : U_{n-1} \rightarrow U_n$ and it follows that there is a nonempty interval $U' \subseteq U_1$ such that $T^{n-1}(U') = T|_{U_{n-1}} \circ T|_{U_{n-2}} \circ \dots \circ T|_{U_1}(U') = U_n$. Since $T^{n-1}|_{U'}$ is continuous, there exists $x \in V_1$ such that $T^{n-1}(x) = x$. \square

REFERENCES

- [1] D. E. ATKINS, *Higher-radix division using estimates of the divisor and partial remainders*, IEEE Trans. Comput., C-17 (1968), pp. 925–934.
- [2] R. BOWEN, *Bernoulli maps of the interval*, Israel J. Math., 28 (1977), pp. 161–168.
- [3] J. COCKE AND D. W. SWEENEY, *High Speed Arithmetic in a Parallel Device*, Tech. report, IBM, February 1957.
- [4] R. L. DEVANEY, *An Introduction to Chaotic Dynamical Systems*, 2nd ed., Westview Press, Boulder, CO, 2003.
- [5] C. V. FREIMAN, *Statistical analysis of certain binary division algorithms*, Proc. IRE, 49 (1961), pp. 91–103.
- [6] A. N. KOLMOGOROV, *A new metric invariant for transient dynamical systems and automorphisms in Lebesgue spaces*, Dokl. Akad. Nauk SSSR (N.S.), 119 (1958), pp. 861–864 (in Russian).
- [7] A. N. KOLMOGOROV, *Entropy per unit time as a metric invariant of automorphisms*, Dokl. Akad. Nauk SSSR, 124 (1959), pp. 754–755 (in Russian).
- [8] A. LASOTA AND M. C. MACKAY, *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*, 2nd ed., Springer-Verlag, New York, 1994.
- [9] A. LASOTA AND J. A. YORKE, *On the existence of invariant measures for piecewise monotonic transformations*, Trans. Amer. Math. Soc., 186 (1973/1974), pp. 481–488.
- [10] F. LEDRAPPIER, *Some properties of absolutely continuous invariant measures on an interval*, Ergodic Theory Dynamical Systems, 1 (1981), pp. 77–93.
- [11] M. MCCANN AND N. PIPPENGER, *SRT division algorithms as dynamical systems*, in Proceedings of the 16th IEEE Symposium on Computer Arithmetic, 2003, pp. 46–53.
- [12] G. METZE, *A class of binary divisions yielding minimally represented quotients*, IRE Trans. Electron. Comput., EC-11 (1961), pp. 761–764.
- [13] M. NADLER, *A high-speed electronic arithmetic unit for automatic computing machines*, Acta Tech. Prague, 1 (1956), pp. 464–478.
- [14] D. S. ORNSTEIN, *Bernoulli shifts with the same entropy are isomorphic*, Advances in Math., 4 (1970), pp. 337–352.

- [15] D. S. ORNSTEIN, *Ergodic Theory, Randomness, and Dynamical Systems*, James K. Whittemore Lectures in Mathematics given at Yale University, Yale Math. Monographs 5, Yale University Press, New Haven, CT, 1974.
- [16] K. PETERSEN, *Ergodic Theory*, Cambridge University Press, Cambridge, UK, 1983.
- [17] J. E. ROBERTSON, *A new class of digital division methods*, IRE Trans. Electron. Comput., EC-7 (1958), pp. 218–222.
- [18] V. A. ROHLIN, *Exact endomorphisms of a Lebesgue space*, Amer. Math. Soc. Transl., 39 (1964), pp. 1–36.
- [19] P. SHIELDS, *The Theory of Bernoulli Shifts*, Chicago Lectures in Math., The University of Chicago Press, Chicago, IL, London, 1973.
- [20] R. SHIVELY, *Stationary Distribution of Partial Remainders in S-R-T Digital Division*, Ph.D. thesis, University of Illinois, Chicago, 1963.
- [21] K. D. TOCHER, *Techniques of multiplication and division for automatic binary computers*, Quart. J. Mech. Appl. Math., 11 (1958), pp. 364–384.
- [22] P. WALTERS, *An Introduction to Ergodic Theory*, Springer-Verlag, New York, 1982.