A Sideways View of Stromatolites: Complexity Metrics for Stromatolite Laminae

Michael C. Storrie-Lombardi¹ and Stanley M. Awramik²

¹Kinohi Institute, 530 S. Lake Avenue, #117, Pasadena, California 91101 USA Tel: (626) 390-3328 Fax: (626) 432-7484 e-mail: mike@kinohi.org

²Department of Earth Science, Preston Cloud Research Lab, University of California, Santa Barbara, CA 93106 Tel: (805) 893-3830 Fax: (805) 893-2314 e-mail: awramik@geol.ucsb.edu

ABSTRACT

Stromatolites offer a unique window into 3.5 billion years of evolution of the microbial communities that built them within the context of an evolving Earth. Our interest is not in the microbial life or their external matrix as independent entities, but the appearance and evolution of complexity itself within this biogeological system. We adopt the canonical definition of complexity as the emergence and detection of previously unseen properties (structures, functions, information), and we propose that the defining emergent property of stromatolites apparent to the human expert eye is the lamination.

To develop a quantitative complexity metric for stromatolites, we must ask what makes it possible for the human brain to perceive lamination? Our visual system operates optimally as a difference machine rapidly identifying variations in signal intensity and redundancy in neighboring regions. In stromatolites, such differences are detected by first scanning parallel to the growth surface and then placing layers in context by scanning orthogonal to that surface. We propose that the fundamental metric for stromatolite complexity resides in the laminae themselves and that easily measured differences in luminance, variability, and redundancy between alternating laminae is an emergent feature of stromatolite complexity. The metrics calculated for laminae in photomicrographs revealed significant differences between putative biotic/abiotic laminae. The statistical indices calculated can contribute to stromatolite recognition, description, and classification. The indices are easily calculated in the laboratory or in the field on personal computers. We propose that such statistical information metrics be included as a standard component in the description of extant and fossil stromatolites.

Keywords: Stromatolites, complexity, emergent properties, quantitative metrics

INTRODUCTION

Stromatolites [Awramik, 1971; Walter, 1976; Riding, 1990; Riding and Awramik, 2000] are classically viewed as laminated organo-sedimentary structures formed by microbial trapping and binding of sediment or precipitation of dissolved minerals [Awramik et al., 1976; Burne and Moore, 1987]. However, some researchers prefer a descriptive definition leaving biogenicity an open question [Semikhatov et al., 1979; Grotzinger and Knoll, 1999; Brasier, 2006]. The increased attention to the Archean fossil record and how it might assist in the search for evidence of past life elsewhere in the universe have heightened concern for unambiguous signals of biological activity [Awramik and Grey, 2005]. Stromatolites as indicators of microbial activity have fallen on hard times. However, a persuasive case for biogenicity in some 3.45 Ga-old stromatolites can be made when multiple lines of evidence are presented that can only be explained by one hypothesis, which invokes microbial activity [Hofmann et al., 1999; Allwood et al., 2006]. These then would provide our most ancient record of life on Earth and rival or even surpass the reliability of microbial fossils, which are also controversial [Brasier et al., 2002; Schopf et al., 2002; Brasier, 2006]. Most Archean and Proterozoic stromatolites are composed of fine grained material and commonly display sub-mm scale laminations [Walter, 1972; Grey and Corkeron, 1998]. They often lack unambiguous microscopic evidence for microbial activity [Grotzinger and Knoll, 1999]. Some abiotic precipitates have been put forward as similar in morphology to selected pre-Phanerozoic sromatolites and modeled using simple accretion rules and common diffusion [Sumner and Grotzinger, 1996; 2000] equations based on the "random walk" (e.g., Brownian motion) of particles without invoking biological processes [Grotzinger and Rothman, 1996]. However, conical stromatolites with very steep surface angles continue to defy such modeling attempts unless high coefficients of

> Instruments, Methods, and Missions for Astrobiology IX, edited by Richard B. Hoover, Gilbert V. Levin, Alexei Yu. Rozanov, Proc. of SPIE Vol. 6309, 63090P, (2006) · 0277-786X/06/\$15 · doi: 10.1117/12.679869



Figure 1. Schematic of data flow for human expert and stochastic artificial neural network classification of complex geobiological structures, in this case, stromatolites.

friction are invoked equivalent to those we employ to model organic molecules [*Hofmann et al.*, 1999; *Jogi and Runnegar*, 2003]. All of these issues are important both to paleobiological and exobiological exploration.

Regardless of origin, the recognition of stromatolites is not a matter of computer modeling, but a function of neurobiological human discrimination of a complex structure against a background geological matrix. Complexity, like beauty, resides in the eve of the beholder. Specifically, the recognition of complexity in physics, chemistry, biology, art, and music depends on the interaction between current perceptions and the associative memory skills of a human

expert. The core function of theoretical and experimental science is then to identify a complexity metric capable of measuring the characteristics of the phenomena that capture expert attention. Such a metric is of quite practical importance. For example, remote, automated detection of differences in relative complexity would be of considerable utility for exploratory field work in dangerous extreme environments on Earth and for robotic exploration of Mars or the icy moons of the outer planets. The search for such complexity metrics has been the focus of numerous investigations [Schrodinger, 1943; Chaitin, 1970; Gell-Mann, 1994; 1995]. Techniques proposed to date include estimation of information content [Shannon, 1948; Shannon and Weaver, 1949], induction by algorithmic probabilities [Solomonoff, 1964; 1964], algorithmic information complexity [Chaitin, 1974; Chaitin, 1987; Zurek, 1989], cellular automata [Wolfram, 1984; Mitchell et al.], Boolean concept learning [Feldman, 2001], adaptive diversity [Huberman and Hogg, 1986], diffusion-compression [Allegrini et al., 2003], complexity evolution [McMullin, 2000; Lenski et al., 2003], physical complexity [Adami and Cerf, 2000], neuronal stimulus encoding [Strong et al., 1998; Krahe et al., 2002], statistical complexity [Crutchfield and Young, 1989], and effective complexity [Gell-Mann and Lloyd, 1996]. The difficulties encountered by these efforts can be best exemplified by considering the first and most famous attempt: Shannon information content or Shannon entropy. Shannon's 1948 paper crystallized communication theory by defining three fundamental aspects of any communication: a transmitter, a channel, and a receiver. To Shannon the defining feature of communication is the uncertainty a receiver has about which message might be sent by the transmitter. That uncertainty is the information content or the entropy of the communication. Imagine that a transmitter can only send one message. In that case, the receiver of the message has no uncertainty about which message he can expect. In Shannon's theory the uncertainty is zero. If the transmitter can send either of two messages, the receiver's uncertainty (entropy, information) is 1 in 2 or 50%; for a transmitter equally capable of sending any one of three messages, 33%, and so on. In this system increased complexity is associated with increasing uncertainty or the information possibilities of the transmitter.

But in such a metric there is nothing to distinguish a signal sent by a random number generator from a maximally encoded message in which every number refers to a look-up table with pointers to a set of paragraphs, chapters, and images sufficient to construct a highly complex nuclear power plant. The latter message is of course possible only if the code employed by the transmitter is known to the receiver. If the code is unknown then the phenomena will be indistinguishable from a completely random event. In science we are usually in the position of not knowing the code. We propose that the fundamental recognition of complexity is a human neurological event, an emergent phenomena in its own right that occurs during classification of a message. An expert observer identifies phenomena as interesting or complex by comparing a current message to vivid memories of past events where the

message may be an image, a spectrum, a sound, or written language. In general, what the expert notices and identifies as interesting, complex, emergent features are most often statistically unexpected redundancies against a more random background matrix, or conversely, the unexpected appearance of seemingly random noise against a less dynamic, more predictable, redundant background. In essence, human recognition of interesting, novel, emergent phenomena depends not on the appearance of some magical complexity maximum, but on detection of coresident random and redundant events by an expert observer, and comparison of the immediate data to a massive internal data set acquired across years of study. Figure 1 depicts the data flow that occurs in both a human expert and a stochastic artificial neural network (SANNs). A series of studies into the classification of galaxies, stars, fossils, and biotic/abiotic alteration of deep ocean basalts have demonstrated that SANNs can be trained by human experts to reproduce their classification decisions. The interested reader is referred to the cited literature for a full discussion of the stochastic algorithmic approach [*Storrie-Lombardi et al.*, 1992; *Storrie-Lombardi et al.*, 1994; *VonHippel et al.*, 1994; *Lahav et al.*, 1995; *Storrie-Lombardi et al.*, 2004; *Storrie-Lombardi and Fisk*, 2004; *Storrie-Lombardi and Hoover*, 2004; *Storrie-Lombardi and Hoover*, 2005]. The critical item in all of these studies has not been the algorithm, but the appropriate selection of measurements most likely to capture the same emergent phenomena detected by the human expert(s).

In the present work we propose that the emergent phenomena to be measured in stromatolites are the light and dark laminae. We also propose that the laminae characteristics most likely to capture the attention of the expert eye are the fundamental pixel statistics accessible in a digital camera image. Mammalian neurobiology functions optimally when asked to detect differences. From first principles defining a difference machine, we characterize a biogeological structure as a layered phenomenon when there exists at least two lamina classes demonstrating optically detectable statistical differences in the grain size and/or composition. We specifically focus on intra-lamina luminance data captured parallel to the stromatolite growth surface, rather than light-dark luminance alterations in the direction of growth such as lamina width or frequency. Although we will explore inter-laminae interactions briefly, we suspect these changes in lamina width and character across stromatolite growth history, the focus of much speculation over the last thirty years [*Park*, 1976; *Andres and Reid*, 2006], actual reflect a mixture of



Figure 2. Low resolution experiment map for *Conophyton garganicum*.

geological, biological, ecological, and evolutionary variables that will remain difficult to deconvolve in the near term. By focusing on characterization of intra-laminal luminal phenomena we constrict our field of inquiry to the most fundamental stromatolite feature laminae. We must stress that the laminae optical differences noted by human eye or recorded by a digital camera are a function of grain shape and size, arrangement of the grains, and chemical composition. In this work we do not attempt to deconvolve chemical from structural events. We note in passing, however, that the data analysis strategy employed here has been previously used to merge structural and chemical data sets in other fields of inquiry including the classification of galaxies, stars, and basalt weathering (ibid).

The visual system has evolved to first register differences in luminance (brightness), then variance, and finally periodic vs random fluctuation. Differences in structure and chemical composition can be captured using first order statistics of image pixel grayscale values and relative lossless compression of matching rectangular images of each lamina class and the intervening boundary zone. Clearly, a major variable affecting the interpretation of detectable differences in illumination and relative compressibility indices in stromatolites is diagenesis. Untangling

complex biosignatures comprised in part of (a) chemical and structural alterations occurring during the lifetime of the microbial community, (b) the components of the signature that are the result of changing chemical conditions associated with degradation of the microbial mat (e.g., Decho and Kawaguchi, 2003) and (c) diagenetic alteration will require a series of laboratory simulations describing the impact of these phenomena. In the current work we will not attempt that task, but will focus on outlining the components of a reasonable statistical description of stromatolite laminae using a single example, *Conophyton garganicum*. We describe (a) the stromatolite and (b) the empirical complexity estimation method. We then present preliminary findings supporting the hypothesis that stromatolite laminae characteristics, including diagenetic alteration, can be easily and reliably described using the fundamental statistics available in a simple RGB digital image and an easily implemented compression metric. This paper begins to explore how the methods we have developed here might potentially be employed as a biogenicity index. Based on such observations as the Gunflint Formation's alternating dark, microfossil-rich and light, microfossil-poor laminae (see Awramik and Semikhatov, 1979), we have postulated that dark laminae were under greater microbial influence than the light laminae. Laminae may represent an alternation between microbial-rich and the biota-active accretion of sediment (dark lamina) and biotically less active, passive (possibly abiotic) accumulation of sediment and/or precipitation of mineral matter (light lamina).

MATERIALS

To illustrate the methodology we present an analysis of silicified *Conophyton garganicum* Korolyuk (emend. Komar, Raaben, and Semikhatov) from the Paleoproterozoic Amelia Dolomite collected by Preston Cloud in 1965 in the Northern Territory (NT), Australia [*Cloud and Semikhatov*, 1969]. *C. garganicum* is a cylindrical, columnar-layered stromatolite with conical laminae, pronounced axial zone, macrolaminae, and a striated microstructure [*Korolyuk*, 1963; *Komar et al.*, 1965]. Examples with diameter up to 1.5 meters and a synoptic relief (height of column when forming) of up to 15 meters are known [*Komar et al.*, 1965]. The sample chosen from the Amelia Dolomite is columnar-layered with an axial zone, millimeter-scale dark and light macrolaminae, and submillimeter-scale laminae within the macrolaminae (Figure 2). A wide range of diagenetic alteration is easily identified at millimeter and sub-millimeter scales (Figure 3). The Amelia Dolomite is a stromatolite-rich unit, 150 to 180 meters thick, that crops out in the NE portion of the NT, near the Gulf of Carpenteria [*Jackson et al.*, 1987]. The age of the formation is ~1650 Ma [*Page et al.*, 2000].

METHODS

Image Acquisition – Images were acquired by photographing, 600W base-illuminated 56 x 75 mm thin sections using Nikon D70S and D200 digital SLR cameras equipped with an AF Micro Nikkor 60 mm lens. Initial NEF images were converted to uncompressed TIF files.

Lossless Compression Estimators – The relative redundancy and randomness of an optical image can be approximated to first order using lossless compression algorithms [Ziv and Lempel, 1977; Nelson and Gailly 1995]. These algorithms reduce file size by identifying and re-coding redundant information. For instance, if the algorithm detects one thousand "0" characters in an image it does what we would do – it simplifies. Instead of storing a thousand digits, it stores the binary equivalent of "start at position x|y and insert '0' x 1000." Now it has stored a few dozen characters instead of 1000. So, a simple complexity index would be C= compressed/raw file size. Note that decompression of the compressed information will return the original data completely unaltered. As pointed out above, the compressibility-incompressibility distribution and the proper way to define the interaction between those two phenomena in a wide variety of natural systems are not yet known making the calculation of an absolute complexity metric beyond the reach of our current science. As a pragmatic alternative, earlier work implemented a matched-pair strategy to estimate relative compressibility by comparing an object to a surrounding environmental matrix [Corsetti and Storrie-Lombardi, 2003; Storrie-Lombardi et al., 2003; Storrie-Lombardi et al., 2004]. In these studies the differential compressibility index, δ_c , compared the lossless compressibility of the image in a target area, C_i , to the compressibility of the surrounding matrix, C_m , where

$$\delta_{\rm c} = \log[1 + (C_t - C_m) / (C_t + C_m)].$$

The absolute value of δ_c provided the magnitude of the difference in complexity independent of whether the target was more or less compressible than the surrounding matrix. While this is a mathematically simple and elegant

methodology, unfortunately, ratios hide the most fundamental question about a difference between two phenomena: is a shift due to an increase in one phenomena or a decrease in the second? In our current state of ignorance about the proper mathematical representation of emergent phenomena, throwing any data away must count as a high-risk, low-gain behavior. Ratios will only be of real utility when there is a much larger and more heterogeneous reference database of well-characterized biogeological targets. After employing the lossless compression technique to estimate redundancy in subocean basalt [*Storrie-Lombardi and Fisk*, 2004], meteorite [*Storrie-Lombardi and Hoover*, 2005], biological [*Storrie-Lombardi*, 2006], and paleobiological [*Storrie-Lombardi and Brown*, 2004; *Storrie-Lombardi et al.*, 2004] targets, it now appears more intuitive and more informative to report raw compression fractions for each target class. A high-compression index means the target is less compressible, contains less redundant or periodic

phenomena, and, hence, is less predictable. It is extremely important to note that both biotic and abiotic phenomena can exhibit compressible and incompressible features. The important point is to detect differences characterizing specific targets as distinct from one another and/or distinct from a common background. For this work we report the simple fractional compression index for RGB TIF files

$$Cx = F_C / F_R$$

where, F_R is the raw file size as reported in a standard Unix shell window (Mac Os X 10.4) using "ls –l" and F_C is the compressed file-size following use of the standard Unix *gzip* implementation of the Ziv-Lempel compression algorithm.



Figure 3. Experiment map of *Conophyton garganicum* at 8mm/pixel spatial resolution. Photograph taken with horizontal lamina orientation. Direction of growth is from bottom to top of image.

RESULTS

Low Resolution RGB Image Statistics - The defining low-magnification characteristic of the sample depicted in Figure 2 is the appearance of the multiple dark and light macrolaminae. The macrolaminae are 1-1.5 mm thick and are millimeters in length. These dark and light macroscopic regions are then further subdivided into finer laminae with thickness of less than a 0.1-0.2 mm. In addition to this variability in light transmission at the macroscopic level, there is also a wide range in preservation of the fine laminae. Visual examination of low and medium resolution images of *Conophyton garganicum* (Figures 2 and 3) reveals several areas of relatively unaltered dark and light macrolaminae but wide variation in average light transmission for both dark and light laminae. Since the impact of the regional variation in transmission was unknown, 64x64 pixel image samples (approximately 0.5 x 0.5 mm) were acquired from the dark (*D*), boundary (*B*), and light (*L*) regions at each of four distinct sites (regions 1-4 in Figures 2 and 3). We calculated image statistics including mean, minimum, maximum, variance, mode, modal number (N) and determined lossless compressibility indices for a total of 60 images to characterize photon transmission characteristics at 8 μ m/pixel resolution.

Results are described for each of two macroscopic lamina classes identified as "Dark" (D) and "Light" (L) as well as for the "boundary" images (B). Boundaries (edges) are of particular interest because of the prominent role they play in human neural pattern recognition and because they often are characterized by non-linear shifts in redundancy easily detected by lossless compression algorithms as a region of high complexity.

In the first region 15 image samples were acquired from an area exhibiting relatively well-preserved dark and light regions and a sharp boundary region. D, B, and L image samples were acquired from left to right in matched triads along the D|L boundary. Region 2 is an example of severe diagenetic change in a light macroscopic lamina. D, B, and L image triads were obtained from 10 sequential sites starting at the left edge of Region 2. The first 4 sites, L1-L4, exhibit marked loss of matrix secondary to diagenetic change resulting in what appear to be lacunae with thin but prominent boundaries. The last 6 images sampled sites appearing fairly well preserved to the eye. Region 3

contains two neighboring dark and light areas that seem to the eye to blend smoothly into one another without a clear, well-defined boundary. Region 4 contains two areas exhibiting fine, well-defined laminae, but one area is dominated by dark and one by light laminae. The boundary region is a wide discontinuity filled with an almost opaque material. To the eye the region of dark banding exhibits much more well-defined boundaries between the sub-millimeter laminae. Figure 4 is a higher resolution photomicrograph of region 5 (Fig. 2 and 3) to explore individual laminae at a spatial resolution of 2.6 μ m/pixel. This region has been morphologically well-preserved by transformation of carbonate to chert early in the diagenetic process.

Image statistics for the 64x64 pixel (8 μ m/pixel) samples (N=60) are reported by region in Table 1. Mean, minimum, maximum, and modal variables differ significantly for dark and light macrolaminae. Table 2 lists the compression indices statistics for all four regions. As expected, images containing the boundary between the two types of macrolaminae are less compressible than images of either dark or light macrolaminae alone. In Region 1, light macro-laminae appear less compressible than dark in four of the five samples. The one triad with similar compression indices for dark and light laminae sampled may have been compromised by submillimeter diagenetic alteration not apparent to the eye at this magnification.

Table 1. Image (64x64 pixel) statistics (8 μm/pixel)								
Exp	D L	Ν	Mean	Min	Max	Mode	Mode N	Variance
1	D	5	67.6±3.8	39.0±1.6	102.0±6.6	60.4 ± 8.4	193.02±6.2	120.4±46.7
	L	5	108.6±3.2	62.8±14.5	142.6±8.2	111.4±2.4	187.8±39.2	109.6±27.7
2	D	10	78.1±3,1	42.7±9.1	115.3±12.6	78.0±3.4	222.6±25.1	97.7±43.0
	L	10	135.2±9.5	13.0±20.8	78.0±9.3	170.2±12.8	135.0±39.1	156.1±89.1
3	D	10	61.5±5.3	34.4±6.6	96.1±6.1	61.6±7.8	188.2±24.1	107.9±30.5
	L	10	105.6±3.0	56.9±23.2	145.5±8.2	104.1±3.0	200.8±25.4	97.8±26.8
4	D>L	5	70.6±1.5	37.0±4.5	101.2±6.8	70.8±2.9	225.4±48.4	89.4±21.8
	L>D	5	85.0±2.7	51.4±5.8	117.2±11.9	83.8±6.5	226.0±55.3	93.4±50.4

Region 2 shows a gradual increase in laminae damage as we scan from the left to the right. Starting in the left portion of the region, sites 1-5 exhibit marked loss of laminae detail due to diagenetic alteration. Sample sites 6-10 appear fairly well preserved to the eye. Light and dark compression indices are clearly distinguishable from each other and from the transition area in the left of Region 2. The most dramatically altered sample sites to the right of the region exhibit complexity levels similar to boundary area condition. This is not unexpected, since the eye identifies the regions as highly altered by detecting the lacunae boundaries. These boundaries differ significantly from laminar boundaries with the latter being quite linear and continuous as opposed to the curvilinear, swirling, and broken boundaries separating the altered lacunae. However, boundary conditions appear more complex to both the eye and the lossless compression algorithm than either the dark or light unaltered matrix.

Region 3 contains two neighboring sets of dark and light macrolaminae appearing to differ only in relative opacity to the eye. In both regions laminae appear fairly well-preserved and show no evidence of the smoothing and lacunae characterizing the more highly altered area in other regions. The boundary between dark and light appears somewhat blurred to the eye. In spite of this relatively smooth blending the compression algorithm detects a significant difference between the boundary and lighter and darker zones in 8 of 10 samples. The algorithm agrees with the human observation that the dark and light zones are otherwise quite similar and detects no significant difference in complexity indices. Region 4 compares two neighboring zones with fine laminae, one dominated by dark and one by light material. The boundary is a pronounced discontinuity filled with an almost opaque material. Again the images including the boundary are the least compressible, but now there is a reversal in compressibility for light and dark macro-laminae.

Global statistics for 64x64 pixel $8\mu m$ low resolution images – Examination of the correlation and covariance (not shown) matrices for pixel and compression variables reveals strong correlations between mean, minimum, maximum, and mode variables, and between modal N and pixel variance. Principal component analysis extracts three factors accounting for > 90% of the information in the data. Factor 1 is comprised predominantly of the covariant mean, minimum, maximum, and mode variables. Factor 2 is strongly correlated with modal N and shows strong negative correlation with pixel luminance variance. Compression index (*Cx*) is the primary component of

Factor 3 with minimal contribution from other variables. It is important to note that the information provided by the pixel variance metric is orthogonal to that provided by the compression index. While dark and light laminae clearly differ from one another on the basis of mean photon transmission and in many cases may exhibit differences in variance and other distribution parameters, the compression index provides information not contained in or constrained by these factors. These comments focus on mesoscopic observations and experiments that would be accessible to the low magnification cameras commonly included in rover work on Mars or in field kits for extreme environment field work on Earth. However, the marked variability in compression across the four sampling sites emphasizes the necessity of focusing on matched laminae if we wish to control for diagenetic alteration. We next present luminance, variability, and compressibility data characterizing sub-millimeter dark and light laminae.

Table 2: Compression (Cx) statistics for Dark (D), Light (L), and Boundary (B) images (8 µm/pixel)								
Ex	D	В	L	[D-B]	[L-B]	[D-L]		
р								
#	(Coi	mpression mean	$\pm 1\sigma$)	(Paired t-test mean difference, 2-tail significance)				
1	0.601 ± 0.011	0.652 ± 0.018	0.619 ± 0.015	-0.050, p<0.0002	-0.033, p<0.001	-0.017,		
						p<0.017		
2	0.625 ± 0.019	0.690 ± 0.019	0.649 ± 0.019	-0.065, p<0.0001	-0.040, p<0.0025	-0.024,		
						p<0.013		
3	0.613 ± 0.017	0.672 ± 0.021	0.619 ± 0.020	-0.059, p<0.0001	-0.053, p<0.0002	-0.006,		
						p<0.493		
4	0.618 ± 0.009	0.679 ± 0.017	0.591±0.026	-0.060, p<0.0008	-0.088, p<0.0001	0.028, p<0.065		



Figure 4. Photomicrograph of lamination in *Conophyton* garganium showing the good preservation of laminae in the chert, relationships between chert and carbonate, and the complexity of the replacement and diagenesis.

(arrow). The area to the right of D shows a portion that is primarily carbonate that preserved some of the original lamination. However, neomorphism has also produced some areas of larger carbonate crystals (arrows). Note the transition of the carbonate-dominated lower portion to the right of D to chert that has replaced the carbonate earlier in diagenesis, preserving some of the finer details of lamination that were lost during post-chert diagenesis of the carbonate layer. 190 x 4 pixel (0.5 x 0.01 μ m) images were obtained at 2.6 mm/pixel spatial resolution from dark and light laminae as matched pairs (N_{dark} = 23, N_{light} = 23) from well-preserved, chert-

Submillimeter luminance, variability, and compressibility – Conophyton in the Amelia Dolomite is silicified. [Jackson et al., 1987] The chert has preserved fine details of the lamination (Figure 4), which make our study possible. The area to the right of A clearly shows that chert replaced laminated carbonate relatively early during diagenesis. Laminae in carbonate can be traced into the chert (arrow). Subsequent to chert replacement, the carbonate underwent further crystallization obscuring much of the fine details of the lamination. The chert now preserves laminae more faithfully than the adjacent carbonate, which is coarser grained than the chert. The area to the right of B shows good preservation of lamina in chert with the post-chert recrystallization of remaining carbonate obscuring original lamination. The carbonate to the right of C has undergone even more diagenesis with neomorphism forming larger crystals of carbonate

rich sites B and B' (Figure 4). The image luminance and compression statistics appear in Table 3. Dark and light laminae were easily distinguished at 23 of 23 sites by both compression and luminance statistics.

Table 3. Image (190x4 pixels) statistics for dark and light laminae.										
	Cx	L[mean]	L[min]	L[max]	L[mode]	[modeN]	L[var]	FK ¹	P>K²	BKS ³
Dark	0.63 ± 0.04	45.7±6.6	29±5	81±19	42±6.6	55±20	102±67	35±12	2.3±9.8E-09	0.82±0.04
Light	0.76 ± 0.02	90.9±15.0	50±11	156±20	79±19	24±5	546±237	34±13	4.3±18.0E-0	80.81±0.04
t-test	4.73E-13	5.09E-13	3.26E-8	1.29E-9	5.20E-9	1.05E-7	1.78E-8	0.726	0.291	0.172
White Noise Periodicity Statistics: [1] Fisher's Kappa; [2] Probability>Kappa; [3] Bartlett's Kolmogorov-Smirnov										

The dark-light laminae distinction is accomplished in 23 of 23 pairs by either Cx or mean luminance and luminance variance correctly identifies light and dark lamina in 22 of 23 pairs. Periodicity estimates including the Fisher's Kappa, Probability>Kappa, and Bartlett's Kolmogorov-Smirnov statistics detect wide variability in periodicity for both light and dark laminae at 2.6 μ m/pixel resolution. Figure 5 depicts the results of first producing an unsupervised, autonomous classification of the 23 pairs of laminae using hierarchical cluster analysis (HCA). Artificial neural network (ANN) prediction of the correctness of the HCA classification using four variables [lossless compression (Cx), mean luminance (Lmean), luminance variance (Lvar), and Bartlett's Kolmogorov-Smirnov estimate of periodicity (BKS). PCA detects strong correlation between 3 of the variables (Cx, Lmean, Lvar) and collapses them into first principal component accounting for 67% of the information in the data set. PCA factor 2 comprised predominantly of the BKS periodicity estimate contributes 26% of the data set information. Periodicity does not distinguish between light and dark lamina due to the extreme variance depicted by the 1 σ error bars in spectral density plot. Examination of photomicrograph reveals marked intrusion of dark lamina into all light lamina. Multiple dark-light edge transitions accompanying these intrusions produces the low compressibility index (Cx).

DISCUSSION

The metrics we have presented here for macroscopic and microscopic laminar phenomena apparent in stromatolite photomicrographs are easily generated using readily available digital camera and laptop computer technology. The metrics quantify redundant and random phenomena in a complex biogeological matrix. We propose that the data critical for quantifying human expert identification of stromatolite complexity can be found in the intralamina aperiodic and periodic alterations in luminance in the dark and light laminae and that the easily measured variations between alternating laminae is an emergent feature of stromatolite complexity. The luminance and compression metrics calculated for laminae in photomicrographs revealed significant differences between dark and light laminae and can serve as an automated edge-detection methodology. We postulate that the initial passive, abiotic sedimentation in the absence of biotic alteration would have left a geological matrix characterized by relatively random distributions of particle sizes and shapes. Photomicrographs at a scale capable of resolving those particles would exhibit minimal redundancy and be relatively incompressible. The relatively smooth textures in the dark laminae in Conophyton exhibit significant redundancy and are more easily compressed. We speculate that the diminished redundancy is the result of microbial alteration, with the laminae representing a temporal alternation between biota-active accretion of sediment (dark) and abiotic passive accumulation of sediment (light). However, testing the validity of that hypothesis is beyond the constraints of the current study. The light laminae in this sample appear to be a mixture of light material with irregular incursions of dark. Such a dynamic system produces a high degree of uncertainty or randomness into the photomicrograph pixel statistics. The algorithms employed here efficiently detect the differences capturing the interaction as high variance and low compressibility.

It appears the indices calculated can contribute to quantification and automation of stromatolite recognition, description, and classification. Boundaries are easily distinguished greatly simplifying the pattern recognition task for robotic devices. The metrics proposed also facilitate detection and analysis (or avoidance) of regions exhibiting diagenetic alteration. In particular we point out that it is important to distinguish compression differences arising from differences in light transmission homogeneity in laminae versus higher order phenomena such as periodic alterations in accretion or subsequent diagenesis. We suspect the detection of true periodic events may evolve into an excellent probe for tracking temporal variations in initial formation and subsequent diagenesis.



Figure 5. Neural network (ANN) prediction of HCA classification using four variables: lossless compression (Cx), mean luminance (Lmean), luminance variance (Lvar), and Bartlett's Kolmogorov-Smirnov estimate of periodicity (BKS). PCA detects strong correlation between 3 of the variables (Cx, Lmean, Lvar) and collapses them into first principal component. Periodicity does not distinguish between light and dark lamina due to extreme variance depicted by the 1σ error bars in spectral density plot. Examination of photomicrograph reveals marked intrusion of dark lamina material into each light lamina. Multiple dark-light edge transitions produce low compressibility index (Cx).

Even though the global statistics for systematic appearance of periodic phenomena (Table 3) show insignificant differences between dark and light laminae, individual analysis of spectral densities for each pair reveals that while some laminae exhibit patterns indistinguishable from white noise, others contain strong periodic signals repeating approximately every 10-30 μ m. Figure 6 depicts the mean spectral density for the 23 pairs of light and dark laminae. The periodic signal is strong enough to make an appearance in the mean density plot for light laminae, but remains buried in the sample variance (error bars are one sigma).

We propose that the human expert identifies significant complexity in an outcrop or an image of a fossil stromatolite by comparing fundamental luminance difference in a specific sample with a large set of images filed in visual memory centers over decades of field and laboratory observations. Estimation by the human expert of the true complexity apparent in a specific stromatolite sample can only be accomplished by placing the observed object in a broader context. If we are to look forward to a time when the subjective evaluations of individual human experts can be codified into algorithms and instrumentation for remote exploration of other terrestrial bodies, we need to start the process of building an easily accessed database containing the same morphological and chemical information employed by the human observer. We recommend that information metrics such as the fundamental image luminance statistics and Ziv-Lempel lossless compression estimates presented here be included as a standard component in the published descriptions of stromatolites. Future work in our laboratories will focus on delineating the statistics characterizing self-organizing, sub-millimeter-scale laminae and training stochastic probabilistic neural networks to mimic the classification decisions of a human expert.

ACKNOWLEDGEMENTS

This work was funded in part by the Kinohi Institute. MCSL would like to thank Clare College, Cambridge University and the Harrison Watson foundation for ongoing encouragement and support of this series of investigations into stochastic classification. MCSL's initial research on stromatolite complexity was supported by Caltech/JPL's Director's Research Discretionary Fund. SMA's initial research on stromatolite laminae was supported by NASA's Exobiology Program.

REFERENCES

Adami, C. and N. J. Cerf (2000) Physical complexity of symbolic sequences. Physica D 137, 62-69.

- Allegrini, P., V. Benci, P. Grigolini, P. Hamilton, M. Ignaccolo, G. Menconi, L. Palatella, G. Raffaelli, N. Scafetta, M. Virgilio, and J. Yang (2003) Compression and diffusion: a joint approach to detect complexity. *Chaos, Solitons & Fractals 15*, 517-535.
- Allwood, A. C., M. R. Walter, B. S. Kamber, C. P. Marshall, and I. W. Burch (2006) Stromatolite reef from the early Archaean era of Australia. *Nature 441*, 714–718.
- Andres, M. S. and R. P. Reid (2006) Growth morphologies of modern marine stromatolites: A case study from Highborne Cay, Bahamas. *Sed. Geol.* 185, 319-328.
- Awramik, S. M. (1971) Precambrian columnar stromatolite diversity reflection of metazoan appearance. *Science 174*, 825.
- Awramik, S. M. and K. Grey (2005) Stromatolites: Biogenicity, biosignatures, and bioconfusion. *Proc. SPIE 5906*, 59060P1–5906P9; doi:10.1117/12.625556.
- Awramik, S. M., L. Margulis, and E. S. Barghoorn, 1976: Evolutionary processes in the formation of stromatolites. *Stromatolites*, M.R.Walter, Ed., Elsevier, 149-162.
- Brasier, M., McLoughlin, N., Green, O., and Wacey, D. (2006) A fresh look at the fossil evidence for early Archaean cellular life. *Phil. Trans. R. Soc. B* 361, 887–902.
- Brasier, M. D., O. R. Green, A. P. Jephcoat, A. K. Kleppe, V. Kranendonk, L. M. J., J. F., A. Steele, and N. V. Grassineau (2002) Questioning the evidence for Earth's oldest fossils. *Nature 416*, 76-81.
- Burne, R. V. and L. S. Moore (1987) Microbialites: organosedimentary deposits of benthic microbial communities. *Palaios 2*, 241-254.
- Chaitin, G. J. (1970) To a mathematical definition of "Life". ACM SICACT News 4, 12-18.
- Chaitin, G. J. (1974) Information-theoretic computational complexity. *IEEE Transactions on Information Theory IT-*20, 10-15.
- Chaitin, G. J., 1987: Algorithmic Information Theory. Cambridge University Press.

- Cloud, P. E., Jr. and M. A. Semikhatov (1969) Proterozoic stromatolite zonation. *Amer. Jour. of Science 267*, 1017-1061.
- Corsetti, F. A. and M. C. Storrie-Lombardi (2003) Lossless Compression of Stromatolite Images: A Biogenicity Index? *Astrobiology* 3, 649-655.
- Crutchfield, J. P. and K. Young (1989) Inferring Statistical Complexity. *Physical Review Letters* 63, 105-108.
- Feldman, J. (2001) Minimization of Boolean complexity in human concept learning. Nature 407, 630-633.
- Gell-Mann, M., 1994: The Quark and the Jaguar. W. H. Freedman and Co.
- Gell-Mann, M. (1995) What is complexity? Complexity 1, 16-19.
- Gell-Mann, M. and S. Lloyd (1996) Information measures, effective complexity and total information content. *Complexity 2*, 44-52.
- Grey, K. and M. Corkeron (1998) Late Neoproterozoic stromatolites in glacigenic successions of the Kimberley region, Western Australia: Evidence for a younger Marinoan glaciation. *Precambrian Research* 92, 65-87.
- Grotzinger, J. P. and D. H. Rothman (1996) An abiotic model for stromatolite morphogenesis. Nature 383, 423-425.
- Grotzinger, J. P. and A. H. Knoll (1999) Stromatolites in Precambrian carbonates; evolutionary mileposts or environmental dipsticks? *Annual Review of Earth and Planetary Sciences* 27, 313-358.
- Hofmann, H. J., K. Grey, A. H. Hickman, and R. I. Thorpe (1999) Origin of 3.45 Ga coniform stromatolites in Warrawoona Group, Western Australia. *Geological Society of America Bulletin 111*, 1256-1262.
- Huberman, B. A. and T. Hogg (1986) Complexity and adaptation. Physica D 22, 376-384.
- Jackson, M. J., M. D. Muir, and K. A. Plumb (1987) Geology of the southern McArthus Basin, Northern Territory. BMR Bulletin 220, 1-173.
- Jackson, M. J., M. D. Muir, and K. A. Plumb (1987) Geology of the southern McArthur Basin, Northern Territory. Bureau of Mineral Resources Bulletin 220, 173.
- Jogi, P. and B. Runnegar (2003) Theoretical support for biological activity in ancient stromatolites. *Astrobiology 2*, 616.
- Komar, V. A., M. E. Raaben, and M. A. Semikhatov (1965) Conophytons in the Riphean of the U.S.S.R. and their stratigraphic significance. *Trudy geologicheskogo instituta, Akademiya nauk SSSR 131*, 72.
- Korolyuk, I. K., 1963: Stromatolites of the Upper Precambrian, Stratigraphy of the USSR, Upper Precambrian. Gosgeoltechnzdat, 479-498 pp.
- Krahe, R., G. Krelman, F. Gabbiani, C. Koch, and W. Metzner (2002) Stimulus encoding and feature extraction by multiple sensory neurons. *Journal of Neurosciences 22*.
- Lahav, O., A. Naim, R. J. Buta, H. G. Corwin, G. de Vaucouleurs, A. Dressler, J. P. Huchra, S. van den Bergh, S. Raychaudhury, L. Sodre Jr., and M. C. Storrie-Lombardi (1995) Galaxies, human eyes and artificial neural networks. *Science 267*.
- Lenski, R. E., C. Ofria, R. T. Pennock, and C. Adami (2003) The evolutionary origin of complex features. *Nature* 423.
- McMullin, B. (2000) John von Neumann and the Evolutionary Growth of Complexity: Looking Backwards, Looking Forwards. *Artificial Life* 6, 347-361.
- Mitchell, M., J. P. Crutchfield, and P. T. Hraber (1994) Evolving cellular automata to perform computations: mechanisms and impediments. *Physica D* 75, 361-391.
- Nelson, M. and J. Gailly 1995: The Data Compression Book. 2nd edition ed. M&T Books.
- Page, R. W., M. J. Jackson, and A. A. Krassay (2000) Constraining sequence stratigraphy in north Australian basins: SHRIMP U–Pb zircon geochronology between Mt Isa and McArthur River. *Australian J. of Earth Sciences* 47, 431-459.
- Park, R. (1976) A note on the significance of lamination in stromatolites. Sedimentology 23, 379-393.
- Riding, R. E., Ed., 1990: Calcareous Algae and Stromatolites. Springer-Verlag.
- Riding, R. E. and S. M. Awramik, Eds., 2000: Microbial Sediments. Springer-Verlag.
- Schopf, J. W., A. B. Kudryavtsev, D. G. Agresti, T. J. Wdowiak, and A. D. Czaja (2002) Laser-Raman imagery of Earth's earliest fossils. *Nature 416*, 73–76.
- Schrodinger, E., 1943: What is Life? Mind and Matter. Cambridge University Press.
- Semikhatov, S. M., C. D. Gebelein, P. Cloud, S. M. Awramik, and W. C. Benmore (1979) Stromatolite morphogenesis - progress and problems. *Canadian J. of Earth Sciences* 16, 992-1015.
- Shannon, C. E. (1948) Mathematical Theory of Communication. *Bell System Tech. Journal* 27, 379-423 and 623-656.

Shannon, C. E. and W. Weaver, 1949: *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, IL.

Solomonoff, R. J. (1964) A Formal Theory of Inductive Inference, Part II. Information and Control 7, 224-254.

Solomonoff, R. J. (1964) A Formal Theory of Inductive Inference, Part I. Information and Control, Part I 7, 1-22.

- Storrie-Lombardi, M. C. (2006) Information content of viruses: impact on long duration space flight. *Kinohi Institute Reports 3*, 5-6.
- Storrie-Lombardi, M. C. and M. R. Fisk (2004) Evidence of biogenic alteration in sub-oceanic basalt glass: complexity image analysis, elemental abundance distributions, and Bayesian probabilistic classification. *Instruments, Methods and Missions for Astrobiology II, (R. B. Hoover, Ed.), Proc. SPIE 5555*, 47-58.
- Storrie-Lombardi, M. C. and R. B. Hoover (2004) Fossil Signatures Using Structural and Chemical Information: Complexity Image Analysis, Elemental Abundance Distributions, and Bayesian Probabilistic Classification. Instruments, Methods and Missions for Astrobiology II (R. B. Hoover, Ed.), Proc. SPIE, 18-30.
- Storrie-Lombardi, M. C. and A. J. Brown (2004) Using complexity analysis to distinguish field images of stromatoloids from surrounding rock matrix in 3.45 Ga Stanley Pool Chert, Western Australia. LPSC XXXV #1414, www.lpi.usra.edu/meetings/lpsc2004/pdf/1414.pdf.
- Storrie-Lombardi, M. C. and R. B. Hoover, 2005: Probabilistic geobiological classification using elemental abundance distributions and lossless image compression in fossils, meteorites, and microorganisms. *Astrobiology and Planetary Missions*, R. B. Hoover, G. V. Levin, and A. Y. Rozanov, Eds., 1-14.
- Storrie-Lombardi, M. C., A. J. Brown, and M. R. Walter (2004) Remote and In Situ Detection of Aqueous and Biotic Alteration: Cyanobacteria in Archean and Modern Australia. *Instruments, Methods and Missions for* Astrobiology II, (R. B. Hoover, Ed.), Proc. SPIE 5555, 270-280.
- Storrie-Lombardi, M. C., O. Lahav, L. Sodre, and L. J. Storrie-Lombardi (1992) Morphological classification of galaxies by artificial neural networks. *Mon. Not. Roy. Astro. Soc*. 259, 8-12.
- Storrie-Lombardi, M. C., M. Irwin, T. von Hippel, and L. J. Storrie-Lombardi (1994) Spectral classification with PCA and artificial neural networks. *Vistas in Astronomy 38*.
- Storrie-Lombardi, M. C., P. Grigolini, S. Galatolo, G. Tinetti, M. Ignaccolo, P. Allegrini, and F. A. Corsetti (2003) Advanced techniques in complexity analysis for the detection of biosignatures in ancient and modern stromatolites. *Astrobiology* 2, 630-631.
- Storrie-Lombardi, M. C., F. A. Corsetti, P. Grigolini, M. Ignaccolo, P. Allegrini, S. Galatolo, and G. Tinetti (2004) Complexity analysis to explore the structure of ancient stromatolites. *Chaos, Solitons & Fractals 20*, 138-145.
- Strong, S. P., R. Koberle, R. R. d. R. van Steveninck, and W. Bialek (1998) Entropy and information in neural spike trains. *Physical Review Letters 80*, 197-200.
- Sumner, D. Y. and J. P. Grotzinger (1996) Were kinetics of Archean calcium carbonate precipitation related to oxygen concentration? *Geology 24*, 119-122.
- Sumner, D. Y. and J. P. Grotzinger, 2000: Late Archean aragonite precipitation; petrography, facies associations, and environmental significance. *Carbonate sedimentation and diagenesis in the evolving Precambrian* world, N. P. James, Ed., SEPM, 123-144.
- VonHippel, T., L. J. Storrie-Lombardi, M. C. Storrie-Lombardi, and M. J. Irwin (1994) Automated classification of stellar spectra: Initial results with artificial neural networks. *Mon. Not. Roy. Astro. Soc. 269*, 97-104.
- Walter, M. R. (1972) Stromatolites and the biostratigraphy of the Australian Precambrian and Cambrian. *Special Papers in Palaeontology 11*, 1-190.
- Walter, M. R., Ed., 1976: Stromatolites. Vol. 20. Developments in Sedimentology, Elsevier, 790 pp.
- Wolfram, S. (1984) Cellular automata as models of complexity. *Nature 311*, 419-424.
- Ziv, J. and A. Lempel (1977) A universal algorithm for sequential data compression. *IEEE Trans. Infor. Theory 23*, 337-342.
- Zurek, W. H. (1989) Thermodynamic cost of computation, algorithmic complexity and the information metric. *Nature 341*, 119-124.