CU STEMinar Journal

Volume 1, Edition 1
February, 2015

Sponsored by the Graduate School at CU Boulder

University of Colorado Boulder

About the STEMinar:

The STEMinar is a graduate student organization at CU Boulder which seeks to promote interdisciplinary interaction among graduate students in STEM departments. The STEMinar hosts seminars given by graduate students about their research. The talks are intended to be relatively non-technical, so that they can be accessible to a broad range of graduate students.

STEMinar speakers have come from a variety of STEM departments including computer science, mathematics, geology, physics, chemistry, evolutionary biology, and several engineering sub-disciplines including aerospace and mechanical. During the Fall 2014 semester, STEMinar talks attracted audience members from at least 13 departments on campus. Moreover, since its inception in August 2013, a majority of the STEMinar talks have been presented by women graduate students, a traditionally under-represented group in STEM disciplines. We hope to continue to find ways to foster diversity and community among the graduate students at CU Boulder.

Thanks to a generous contribution from John Stevenson, the Dean of the CU Boulder Graduate School, the STEMinar has put into place a grant giving program. The individuals who received a $250 STEMinar grant during the Fall 2014 semester were asked to contribute a short document which summarizes their current research project, or reviews a STEM topic that they are interested in writing about. In particular, we have encouraged the authors of these papers to address, if possible, how their research relates to other fields, everyday life, etc.

The STEMinar Journal begins with a list of the talks that were presented during the Fall 2014 semester. The journal will then present the papers that were submitted by our grant recipients during the Fall 2014 semester. We believe that this journal will highlight the ongoing research that is being done by STEM graduate students at CU Boulder.

Title page photo (courtesy of Trubee Davison): A view of the flatirons from South Boulder.

Fall 2014 STEMinar Speakers:

- Noah Williams (Mathematics): Sept. 2, 2014

  Mathematics and the genome rearrangement problem

  Mathematics can be used to help biologists understand evolutionary relationships and diseases such as Neurofibromatosis. In this talk, we will investigate the genome rearrangement problem, whose goal is to find the best way to transform one collection of DNA into another. In particular, we will discuss the Double Cut and Join (DCJ) and Deletion-Insertion models for genome rearrangement and some results pertaining to the distribution of genomes under DCJ distance.

- Osama Bilal (Aerospace Engineering): Sept. 18, 2014

  PhoNonic materials: The mystery of missing frequencies and their applications

  Transmission of everyday sound and heat can be tracked back to a physical particle named "phoNon". Understanding, analyzing and manipulating such particle across multiple scales/disciplines can be achieved using phononic materials. That is a class of material systems composed of a basic building block that repeats in space (i.e. unit cell). Among many features, it exhibits distinct frequency characteristics such as band gaps, where certain frequencies are prohibited from propagation. These properties can lead to tremendous range of applications, ranging from vibration isolation, converting heat-waste into electricity to acoustic cloaking. In this seminar, I will give an introduction to the field, explain the main concepts, potential applications and shed some light on its design and optimization.

- Kristen Brown (Mechanical Engineering): Sept. 30, 2014

  Internalizing air quality and greenhouse gas externalities in the US energy system

  This study provides a unique analysis of how the US energy system might be transformed if energy costs internalized the damages caused by air pollution. The US energy system is modeled including fees on emissions that are based on estimates of the external cost of those emissions. The externalities that are internalized include health and environmental effects as well as climate change. Placing fees on emissions provides more of an incentive to reduce pollution than fees on energy use. This interdisciplinary work incorporates knowledge from engineering, economics, mathematics, medicine, climate research, and atmospheric chemistry.

- Sarah Welsh-Huggins (Civil Engineering): Oct. 16, 2014

  Integrating green and resilient building design for enhanced disaster recovery

  Increased worldwide urbanization in recent years has led to greater vulnerability from seismic and other natural hazards. Integrating performance-based design and assessment of green and hazard- resilient buildings, however, offers alternatives for urban disaster recovery. Performance-based earthquake engineering aims to reduce the dollars, deaths, and downtime associated with a hazard event. Green design, meanwhile, strives to minimize use of natural resources and provide economic gains associated with reduced operating energy or greater occupant productivity. My dissertation research proposes a new life cycle assessment methodology that links the distinct fields of sustainability and hazard-resilience, advancing building design and identifying leverage points for improved disaster mitigation and recovery. The "green-resilience" framework presented here can be used to analyze the implications of disaster mitigation and reconstruction. The methodology quantifies environmental metrics and socio-economic indicators of building performance over its lifetime and accounts for potential risk of hazard events. The utility of the framework will be illustrated with a case study of the design and seismic assessment of a hypothetical commercial building in Los Angeles, CA, U.S.A.

- Helen McCreery (Evolutionary Biology): Oct. 30, 2014

  Collaborating with ants: basic bio, shared lessons, and cooperative transport

  Ant colonies consist of autonomous, simple units capable of remarkable problem solving through cooperation. Ants build massive, complex, underground nests, and many species work together to transport objects hundreds of thousands of times their own weight. How does this complex coordination emerge from groups of simple individuals? This question is inherently interdisciplinary, and there is a long history of collaboration and shared lessons among ant biologists, computer scientists, and mathematicians. For example, algorithms inspired by ants and other social insects are used in swarm robots. In this STEMinar, I will give a brief introduction to ant biology and cooperation, and discuss my interdisciplinary research on cooperative transport.

- Maya Fabrikant (Physics): Nov. 6, 2014

  Frozen molecules and space chemistry

  Cooling molecules allows us to study what kind of chemistry happens at temperatures characteristic of interstellar atmospheres. More practically, cold environments can stabilize very reactive molecules like radicals, which can allow us to study difficult-to-observe reactions. Exotic ultracold molecules like KrB have been created in

the lab with the help of laser cooling and photoassociation, but sufficiently cooling molecules more commonly encountered in nature, like the CH and NH radicals, presents its own unique challenges. I will talk about methods for cooling and detecting these more mundane molecules and what kind of chemistry we might expect to observe.

- Marianne Reddan (Cognitive Psychology): Nov. 19, 2014

  Developing neural signatures for discrete emotional states

  Despite its research popularity and societal importance, emotion remains elusive to neuroscience. Progress is slowed by the field's concern with local representation, a one to one mapping of structure and function. This has led to the identification of many correlates of emotion (e.g. amydala-threat, nucleus accumbens-reward), but the complexity and distributed nature of emotional experiences remain neglected. Our laboratory employs a novel and promising approach to the problem of emotion: We apply machine learning to neuroimaging data in order to identify sensitive and specific signatures of emotional responses. We have successfully developed signatures of pain and negative emotion, which have strong implications for the treatment and diagnosis of emotional disorders. In this talk, I will discuss (1) a ongoing study which aims to classify nuanced emotional experiences, and (2) my work investigating the neural representation of threat-anticipation, and how imagination can change its expression.

- Joey Hubbard (Evolutionary Biology): Dec. 5, 2014

  Title: Genetic and environmental contributions to a divergent plumage trait in barn swallows

  A central theme of evolutionary research is to understand the source and function of phenotypic variation. For most phenotypic traits, sources of variation can be partitioned into genetic or environmental variation. Identifying the proportion of variance due to these components allows for predictions regarding phenotypic responses to variable environments and selection to be made. Barn swallow subspecies have divergent phenotypes and it appears that the focal trait for female preferences has also diverged. We examined the sources of phenotypic variation of melanin-based ventral plumage in two populations of barn swallows: North America (Colorado) and Europe (Czech Republic). In North America, male coloration is the target of sexual selection with darker males achieving higher reproductive success. However, it is unclear what role coloration plays in mate choice decision for our study population in the Czech Republic. In both populations, we found that coloration is explained by both genetic and environmental variation. However, the genetic covariance structure of color traits differed. Together, these results demonstrate that

coloration is influenced by developmental environment more than genetic environment. Thus, in North America, where females prefer males with dark plumage, coloration serves as a better signal of developmental conditions than genetic quality. Moreover, divergent selection on plumage coloration may explain the phenotypic differences among these populations, suggesting a role of sexual selection in the diversification of the barn swallow species complex.

Table of Contents:

PHONONIC MATERIALS: THE MYSTERY OF MISSING FREQUENCIES AND THEIR APPLICATIONS

OSAMA R. BILAL

ABSTRACT. Transmission of everyday sound and heat can be traced back to a physical particle called a “phonon". Understanding, analyzing and manipulating such particle across multiple scales/disciplines can be achieved using phononic materials. That is a class of material systems composed of a basic pattern that repeats spatially. Among many features, it exhibits distinct frequency characteristics such as band gaps, where vibrational waves of certain frequencies are prohibited from propagation. These properties can lead to tremendous range of applications, ranging from vibration isolation and converting waste heat into electricity to more exotic ones like acoustic cloaking. Here I introduce the basic concepts of phononic materials, their classifications and potential applications.

CONTENTS

1. INTRODUCTION

One main aspect of advancement for the human race is measured by the ability to understand, control and manipulate matter. This vindicate the classification of history periods according to the usage of natural material in daily activities (i.e. stone, bronze, iron, modern and computer ages). In order to predict what is next, we should have a closer look at where we stand now. Most of the technological breakthroughs that emerged in the last century were results of our command and control of two main sub-atomic particles; electrons and photons. The manipulation of electrons in semi-conductors enabled the existence of laptops and cell phones, while controlling photons with photonic materials/crystals permitted microwaves and wireless communications. A more recent trend with tremendous potential is related to the utilization of yet another sub-atomic particle, namely the phonon, using *phononic materials* [1]. Numerous studies are investigating their utilization in cloaks for submarines, efficient thermoelectrics, thermally insulated buildings, reduced environmental noise and earthquake protection to name a few [2].

Date: February 18, 2015.
Key words and phrases. Phononic crystals, Acoustic metamaterials.

Phononic materials are periodic systems composed of a basic building block that repeats in space (i.e., unit cell). These materials exhibit distinct frequency characteristics, such as band gaps, frequency ranges where elastic/acoustic waves (i.e., phonons) are prohibited from propagation. Phononic materials can be classified into two subcategories based on the band gap formation mechanism, phononic crystals (PnCs) and acoustic metamaterials (AMs). Wave attenuation in PnCs is a result of Bragg scattering (BS) (i.e., scattering due to periodicity). For AMs, band gaps exist due to the local resonance (LR) phenomena. Owing to the different physics of band gap creation, the ratio between unit cell size and affected wavelength is different for each category. AMs with subwavelength resonance band gaps, can attenuate waves orders of magnitude larger in wavelength than their unit cell size, while PnCs unit cell have to be in the same size as the targeted wavelength.

2. Realization and Modeling

As shown in Fig.1 phononic materials can be modeled, analyzed and realized in different dimensions [3, 4, 5, 6, 7]. Apart from dimensionality, phononic materials are modeled in either continuum (e.g., rod, beam, plate, shell or bulk material) or discrete (e.g., lumped mass system); however, to demonstrate most of its important dynamical characteristics, a simple one-dimensional (1D) spring-mass model is sufficient.

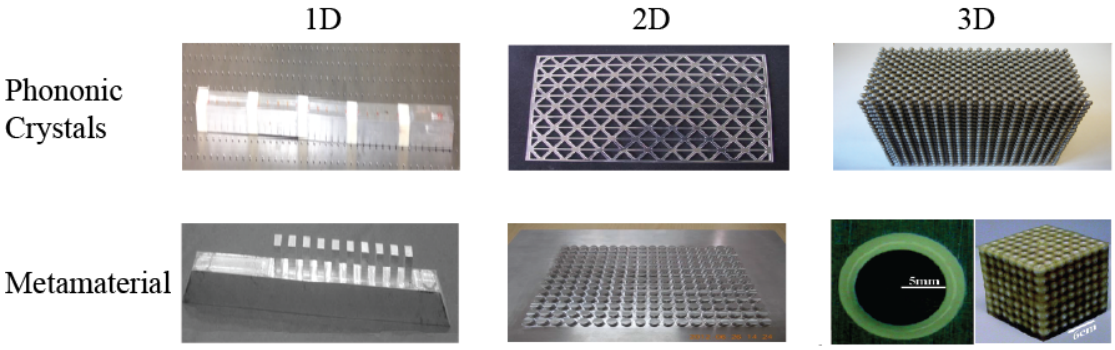


FIGURE 1. Realization of phononic crystals and acoustic metamaterial in one-, two- and three-dimensions

A 1D model of a phononic crystal is represented by two masses m_1 and m_2 connected in series by springs k_1 and k_2 [Fig. 2.a]. The equations of motion for the system, based on Hooke's and Newton's second law, are presented in Eqs. 2.1 and 2.2 for m_1 and m_2, respectively.

$$m_1\ddot{u}_1^j + k_1(u_1^j - u_2^{j-1}) + k_2(u_1^j - u_2^j) = 0 \tag{2.1}$$

$$m_2\ddot{u}_2^j + k_1(u_2^j - u_1^{j+1}) + k_2(u_2^j - u_1^j) = 0 \tag{2.2}$$

where u is the displacement, j is the unit cell index and $\ddot{(.)}$ is the second derivative of $(.)$ w.r.t time.

For the AM, the same number of masses and springs are used, albeit connected differently [Fig. 2.b]. The equations of motion for the system are presented in Eqs. 2.3 and 2.4 for m_1 and m_2, respectively.

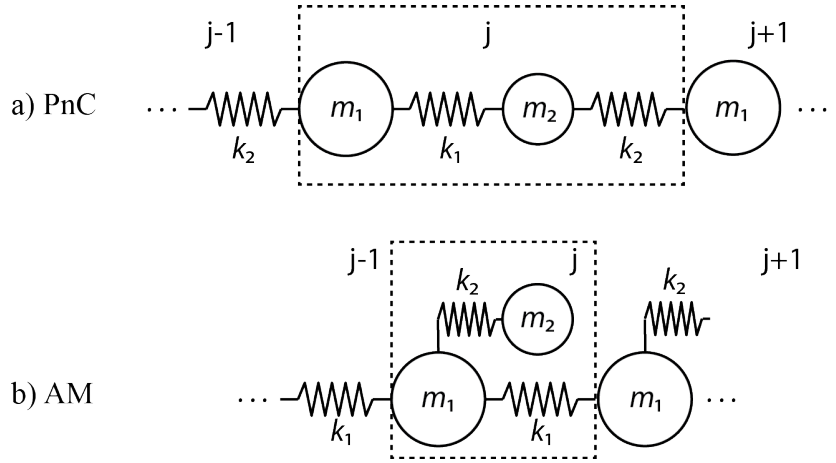


FIGURE 2. simple one-dimensional spring-mass model for a) phononic crystal b) acoustic metamaterial

$$m_1\ddot{u}_1^j + k_1(2u_1^j - u_1^{j-1} - u_1^{j+1}) + k_2(u_1^j - u_2^j) = 0 \tag{2.3}$$

$$m_2\ddot{u}_2^j + k_2(u_2^j - u_1^j) = 0 \tag{2.4}$$

Assuming a generalized Bloch solution of the form $\mathbf{u}^{j+n} = Ue^{i(\omega t + n\kappa a)}$, where ω is the frequency, κ is the wavenumber and a is the unit cell size (i.e. lattice spacing) and grouped in a matrix representation, the equations of motion for both systems yields a complex generalized eigenvalue problem in the form:

$$(\omega^2\mathbf{M} + \mathbf{K}(\kappa))\mathbf{u} = \mathbf{0} \tag{2.5}$$

where mass and stiffness matrices can be written as:

$$\mathbf{M}_{PnC} = \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix}, \mathbf{K}_{PnC}(\kappa) = \begin{bmatrix} k_1 + k_2 & -(k_1e^{-i\kappa a} + k_2) \\ -(k_1e^{i\kappa a} + k_2) & k_1 + k_2 \end{bmatrix} \tag{2.6}$$

$$\mathbf{M}_{AM} = \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix}, \mathbf{K}_{AM}(\kappa) = \begin{bmatrix} 2k_1(1 - \cos(\kappa a)) & -k_2 \\ -k_2 & k_2 \end{bmatrix}. \tag{2.7}$$

3. ANALYSIS

In order to obtain statically equivalent models for PnC and AM, the speed of sound c is matched in the long-wavelength:

$$c = \lim_{\kappa \to 0} \frac{\partial \omega}{\partial \kappa}. \tag{3.1}$$

TABLE 1. Summary of parameters for PnCs and AMs unit cells

Cell	$m_1[kg]$	$m_2[kg]$	$k_1[N/m]$	$k_2[N/m]$	$c(\kappa \to 0)[m/s]$
AM	1	2	10,000	10,000	57.73
PnC	1	2	20,000	20,000	57.73

By using the values in Table 1 for both systems, one can solve the eigenvalue problem in Eq. 2.5 and obtain the dispersion curves of the two material systems that correlate wavenumber to frequency [fig. 3].

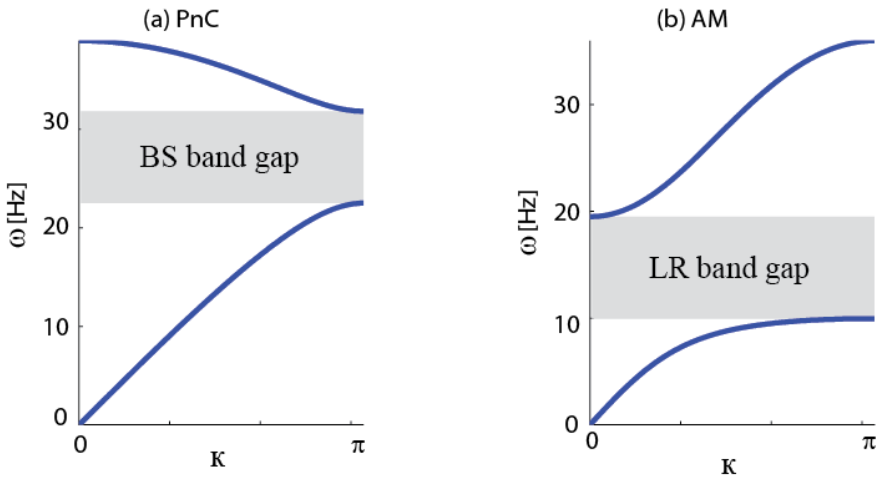


FIGURE 3. Dispersion curves for PnC and AM systems

The two dispersion curves have the same slope (i.e., group velocity) of the first branch for small κ values (i.e., the long-wavelength limit). However, the group velocity decreases quicker in the AM as κ increases, due to the hybridization between the first branch (a.k.a., the acoustic branch) and the local resonance mode. The band gaps are highlighted in gray. The PnC band gap ranges from 22-31 Hz, while for a statically equivalent AM system the band gap exist in a lower frequency range, 10-20 Hz.

4. CONCLUSION

An introduction to phononic materials is presented. After shedding some light on their potential applications,a connection is drawn to electronics and photonic crystals. A distinction between phononic crystals and acoustic metamaterials is made. A simple 1D spring-mass is analyzed and dispersion curves, the key tool in their design and understanding, is presented for each material classification.

REFERENCES

[1] Mahmoud I Hussein, Michael J Leamy, and Massimo Ruzzene. Dynamics of phononic materials and structures: Historical origins, recent progress, and future outlook. *Applied Mechanics Reviews*, 66(4):040802, 2014.

[2] Martin Maldovan. Sound and heat revolutions in phononics. *Nature*, 503(7475):209–217, 2013.

[3] Bruce L Davis, Andrew S Tomchek, Edgar A Flores, Liao Liu, and Mahmoud I Hussein. Analysis of periodicity termination in phononic crystals. In *ASME 2011 International Mechanical Engineering Congress and Exposition*, pages 973–977. American Society of Mechanical Engineers, 2011.

[4] S. Halkjaer, O. Sigmund, and J. S. Jensen. Maximizing band gaps in plate structures. *Struct. Multidiscip. O.*, 32:263–275, 2006.

[5] Hammouche Khales, Abdelkader Hassein-Bey, and Abdelkrim Khelif. Evidence of ultrasonic band gap in aluminum phononic crystal beam. *Journal of Vibration and Acoustics*, 135(4):041007, 2013.

[6] Jin-Chen Hsu, Hua-Shien Hsu, and Tsung-Tsong Wu. Frequency gaps and waveguiding in two-dimensional phononic plates with periodic stepped resonators. In *Ultrasonics Symposium (IUS), 2012 IEEE International*, pages 1750–1753. IEEE, 2012.

[7] ZY Liu, XX Zhang, YW Mao, YY Zhu, ZY Yang, CT Chan, and P Sheng. *Science*, 289(5485):1734–1736, 2000.

E-mail address: osama.bilal@colorado.edu

AEROSPACE ENGINEERING SCIENCES, UNIVERSITY OF COLORADO BOULDER, BOULDER, COLORADO

INTERNALIZING AIR QUALITY AND GREENHOUSE GAS EXTERNALITIES IN THE US ENERGY SYSTEM

KRISTEN E. BROWN

ABSTRACT. The emission of pollutants from energy use has effects on both local air quality and the global climate, but the price of energy does not reflect these externalities. This study aims to analyze the effect that internalizing these externalities in the cost of energy would have on the US energy system. If fees on emissions were in place, producers might be more likely to lower their emissions rates, reducing the negative effects on air quality and in turn on human health and welfare. Consequently, even if fees cause an increase in the price of electricity, the overall cost related to electricity will decrease because external costs will be lowered. When damage costs are included in the price of energy, it is likely that the externalities will decrease, because the increased cost of energy from sources with large external effects will make it less attractive to consumers. Hopefully, such internalization will improve human and environmental health as well as increase the fairness of the market. In this study, we model different policy scenarios in which fees are added to emissions related to generation and use of energy. The fees are based on values of damages estimated in the literature and are applied to upstream and combustion emissions related to electricity generation, industrial energy use, transportation energy use, residential energy use, and commercial energy use. The energy sources and emissions are modeled through 2055 in five-year time steps.

1. INTRODUCTION

Energy use in the US is influenced by many factors, but not all consequences are considered when choosing electricity sources. The emission of pollutants has effects on both local air quality and global climate. The damages from electricity generation in the US in 2005 have been estimated at $62 billion from coal plants and $740 million from natural gas plants [1]. Externalities are activities that affect the well-being of an unrelated group or individual outside the market mechanism. Damages are the monetary value of externalities, such as the value of medical bills from adverse health effects. This study aims to determine how incorporating these damages into energy costs would impact energy use in the US. Damages from criteria pollutants (NO_x, SO_2, particulate matter PM, and O_3) as well as greenhouse gases (GHGs) are accounted for by applying emissions fees equal to the estimated external damages. While many externalities occur due to combustion, other externalities are associated with energy production and use so we consider life cycle emissions.

This paper evaluates how incorporating externality costs into the cost of energy would change energy use and air quality in the United States. A range of damage estimates from the literature is used to construct various policy scenarios, all of which prescribe

Date: January 31, 2015.
Key words and phrases. air quality, energy, externalities, policy.

a set of damage-based emission fees. The policy scenarios will consider damages associated with GHGs and criteria pollutants. The EPA US 9 region MARKAL model will be used to evaluate changes to the US energy system through the year 2055 under these policies compared to a base case with no policies.

2. METHODS

2.1. **Damages.** First, we will discuss criteria pollutant damages. The EPA regulates pollutants considered harmful to public health and the environment. They define a set of six 'criteria' pollutants, carbon monoxide, lead, nitrogen oxides (NO_x), ozone (O_3), particulate matter (PM), and sulfur dioxide (SO_2).[2] The emissions considered here are NO_x, $PM_{2.5}$, PM_{10}, SO_2 and VOC. VOCs are precursor emissions to O_3 and PM. From this point forward, 'criteria pollutant emissions' will refer to emissions of these five air pollutants in order simplify the language. The damages used as fees on criteria pollutants are shown in Table 1 and come from a National Research Council report [1].

pollutant	NO_x	PM_{10}	$PM_{2.5}$	SO_2	VOC
M$/kt	1.97	1.12	21.52	9.75	1.72

TABLE 1. The criteria pollutant damage values used as fees, given in million year 2005 USD/kilotonne.

Global climate change is caused in part by energy related emissions. GHGs are emitted as a result of several uses of energy [3]. There are large discrepancies in the estimated value of damages from climate change. Almost all variation in marginal damage estimates derives from differences in assumptions of discount rate and the magnitude of damages from climate change, especially whether unlikely but catastrophic effects were considered [1]. The greenhouse gas damages used in this study are taken from the Social Cost of Carbon (SCC) used in regulatory impact analysis by the US government [4]. The fees were applied to CO_2 and CH_4. The fees were adjusted for CH_4 using the 100 year global warming potential (GWP) of 28 [5]. These values vary for different years because these species are relatively long lived and so emissions in later years compound with emissions in the present.

	2015	2020	2025	2030	2035	2040	2045	2050
CO_2	35.76	40.46	45.17	48.93	53.64	58.34	62.11	66.81
CH_4	1.00	1.13	1.26	1.37	1.50	1.63	1.74	1.87

TABLE 2. The greenhouse gas damages used as fees, given in year 2005 USD per tonne. The fee in 2055 is the same as in 2050.

2.2. **The MARKAL model.** This study attempts to determine how implementing damage-based emissions fees would change how energy is used in the US in the coming decades. This is done using the MARKAL (MARKet ALlocation) model, which is an energy system model that is used here to compare energy scenarios. MARKAL is chosen because

of its thorough description of the US energy system and the ability to model a variety of fees. The MARKAL model uses linear optimization to determine the least cost solution to fulfill the energy needs of the US from 2005 through 2055 in five year time steps [6]. The model is composed of a set of linear inequalities and assumes a perfectly competitive market so the least cost technology will be used until it has reached maximum capacity and then the second least expensive technology will come into play, and so on. These qualities make it suitable to quantify the effects of policy changes. Constraints can be applied to this model such as maximum allowable emissions representing existing regulations. The model is demand driven, which means all end use demands must be satisfied for every time period in the solution [6]. Generation and conservation technologies are included, so demand can be met partially by using more efficient end-use devices therefore reducing fuel demand. MARKAL considers the economic advantages of different technologies while incorporating external damages as costs ensures that the environmental costs are also considered. The model is further describe in Brown et al. [7]. Additional updates were made to the MARKAL model to better represent the industrial sector and the available control technologies, efficiency improvements and renewable energy sources. Upstream emissions were also added to the model.

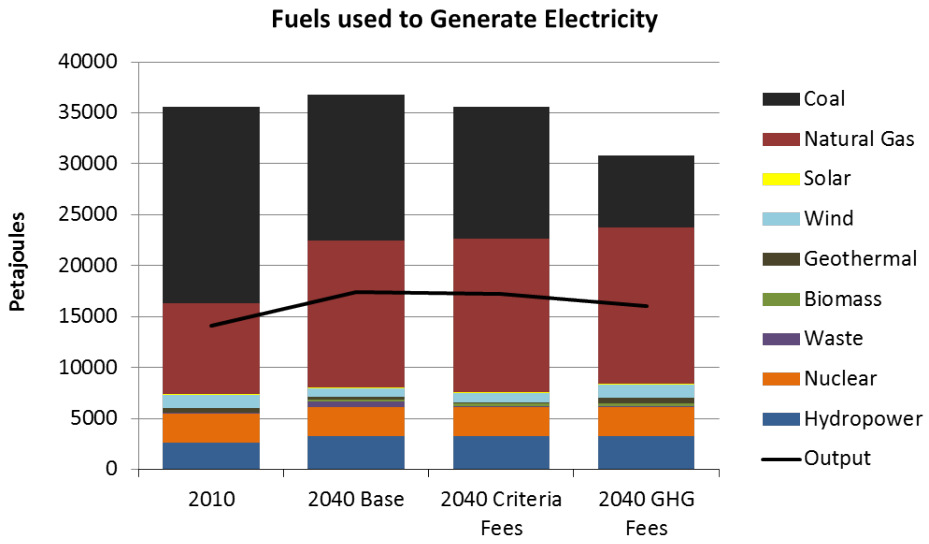
3. RESULTS

When the fees are applied, energy use and emissions are changed relative to the base case. Whether fees are added or not, the energy system will be different in 2040 than it is today. This is due in part to a growing population demanding more energy and in part due to emissions regulations that are already on the books that will require further emissions reductions or more stringent control technologies than are in place today. The figures show results from the base case in 2010 to represent the current energy system. Results are shown from the base case, the criteria pollutant fee case, and the greenhouse gas fee case for 2040 to show the similarities and differences between the policy scenarios after the fees have been in place for 25 years.

Figure 1 shows the fuels used to generate electricity in the US under the different policy scenarios. Electric sector fuel use is similar in 2010 and 2040 in the base case and criteria pollutant fee case. The output of electricity is greater in these cases in 2040 due to an increasing population projected to demand more electricity while the total amount of fuel used is similar due to improvements in efficiency of generation. Natural gas becomes a larger portion of the fuel used to generate electric power in all cases, because older coal fired power plants are retired and new natural gas plants are constructed. In many cases coal is used because the existing coal fired power plants have low cost, but no new coal fired power plants are built. Investments in new natural gas fired power plants are used to replace old coal base-load units due to the lower emissions control requirements associated with a lower emitting fuel and the low cost of natural gas, which makes the total cost of generating electricity over the life of the plant projected to be low. Coal use decreases dramatically with fees, particularly GHG fees. This decrease in coal power is offset by a combination of wind, natural gas, and energy efficiency improvements.

There is relatively little change in the fuels used to generate electricity between the different cases, but that does not mean the policy is not effective. Figure 2 shows the

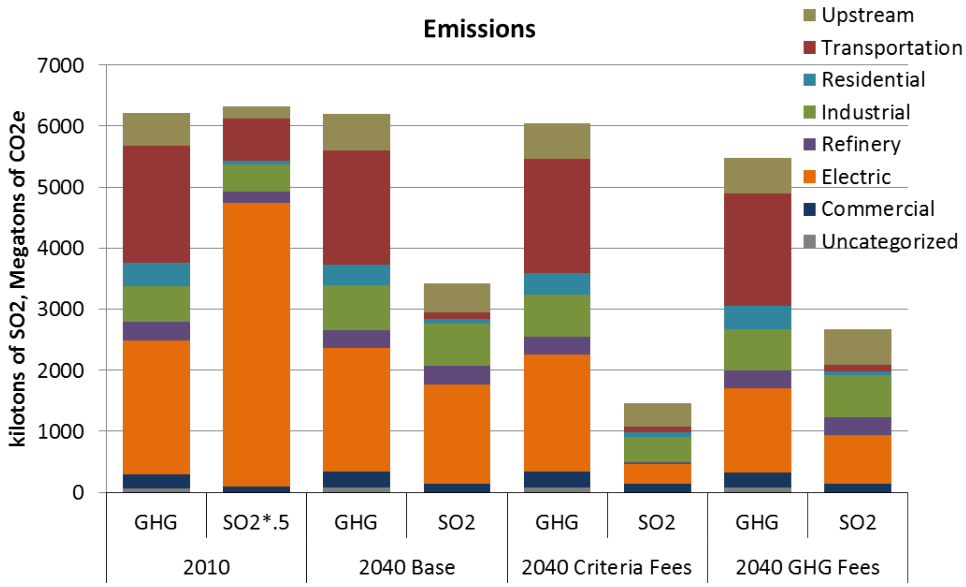
FIGURE 1. The fuels used to generate electricity and the electricity used in different cases, compared to "current" (2010) values.



emissions of representative pollutants in the same cases. SO_2 is shown as an example of a criteria pollutant and the GHG emissions show emissions of CO_2 and CH_4 lumped together as CO_2-equivalent. GHG emissions are relatively unresponsive to criteria pollutant fees, but do decrease when GHG fees are applied. This is because criteria pollutant fees lead to an increase in the use of emissions control technologies. This has little affect on fuel use and the control technologies are typically targeted to remove only the pollutants of interest, i.e. the pollutants with an associated cost. The criteria pollutant emissions are significantly decreased with criteria pollutant fees. This signifies that there are several technologically and economically available emissions control technologies for these pollutants that are not in use because there is no impetus on the industry to apply them. When GHG fees are applied, all emissions decrease and there is a larger affect on fuels used. This is because the control technologies available to reduce GHG emissions are often more expensive than the cost to switch to a different energy source. When switching to a method of energy use with lower combustion emissions, the emissions are reduced indiscriminately, because instead of targeting and removing a particular set of emissions, all emissions associated with the old energy source are removed and replaced by the lower emissions of the new source. Considering these co-benefits, such as a reduction in criteria pollutants in response to climate policy, is an important consideration when writing policy. Another interesting point to note about Figure 2, is that emissions from many energy use sectors are displayed. The use of fuels changed the most in the electric sector, so that sector was shown as an example of fuel use differences, but other emissions reduction methods such as efficiency and control devices were used to reduce emissions in other sectors.

These results can be considered as an indicator of how our energy system might change if damage based emissions fees were implemented. Alternatively, we can consider the resultant emissions levels as a target to consider. Since the emissions were

FIGURE 2. Greenhouse gas (GHG) and criteria polluant (SO_2) emissions from various energy use sectors in different cases, compared to "current" (2010) values. Note that 2010 SO_2 emissions have been multiplied by 0.5 to fit on the scale.



achieved in the model with fees based on damage values, emissions reductions to these levels can be achieved while providing as much health benefit to society as the increased cost of new technologies, given the currently available technologies.

4. ACKNOWLEDGEMENTS

I would like to thank my advisors, Jana Milford and Daven Henze, and my thesis committee, Mike Hannigan, Shelly Miller, and Greg Frost. I would also like to thank the EPA for use of the EPA US 9-region MARKAL database version EPAUS9r_12_v1.1. The results and analysis presented here were derived independent of EPA input. I also thank Nicholas Flores and Garvin Heath for their helpful comments during early stages and Dan Loughlin for assistance in using the EPA MARKAL model. This research was supported by a University of Colorado seed grand from the Renewable and Sustainable Energy Institute and the NASA Applied Science Program NNX11AI54G.

REFERENCES

[1] National Research Council Committee on Health, Environmental, and Other External Costs and Benefits of Energy Production and Consumption, 2010. Hidden Costs of Energy: Unpriced Consequences of Energy Production and Use, Washington, D.C.: The National Academies Press.

[2] US EPA, Office of Air and Radiation, 2015. History of the Clean Air Act. Available at: http://www.epa.gov/air/caa/amendments.html#caa90 [Accessed January 22, 2015].

[3] Ciais, P. et al., 2013. Carbon and Other Biogeochemical Cycles. In T. F. Stocker et al., eds. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment

Report of the Intergovernmental Panel on Climate Change. Cambridge, UK and New York, NY, USA: Cambridge University Press. Available at: http://www.climatechange2013.org/images/report

[4] Interagency Working Group on Social Cost of Carbon, 2013. Technical Support Document: Technical Update of the Social Cost of Carbon for Regulatory Impact Analysis Under Executive Order 12866.

[5] Myhre, G. et al., 2013. Anthropogenic and Natural Radiative Forcing. In H. Zhang et al., eds. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK and New York, NY, USA: Cambridge University Press. Available at: http://www.climatechange2013.org/images/report

[6] Loulou, R., Goldstein, G. Noble, K., 2004. Documentation for the MARKAL Family of Models. Available at: http://www.etsap.org/documentation.asp [Accessed September 23, 2010].

[7] Brown K.E., Henze D.K., Milford J.B., 2013. Accounting for Climate and Air Quality Damages in Future U.S. Electricity Generation Scenarios. Environmental Science & Technology 47, pp. 3065Ð3072.

E-mail address: Kristen.E.Brown@colorado.edu

DEPARTMENT OF MECHANICAL ENGINEERING, UNIVERSITY OF COLORADO, BOULDER, CO

METRICS AND FRACTALS

TRUBEE DAVISON

ABSTRACT. In this paper, we will present a relatively non-technical introduction to the study of fractals. The key ideas that we will discuss include metric spaces, completeness, Lipschitz contractions, measures, compactness, to name a few. These key ideas will be used to describe the three standard methods for constructing the so called fractal set associated to an iterated function system.

1. FRACTALS IN NATURE

In popular culture, it is quite common to hear the word 'fractal' used to describe intricate and self-similar patterns. When we are shopping at the grocery store, for example, we encounter fractal-like objects in the vegetable aisle. The below photograph is a specimen of a *romanesco broccoli* that I purchased last year.

Figure 1: Photograph of a *romanesco broccoli*.

If we observe the *romanesco broccoli*, we see that this vegetable is conical in shape and is comprised of a spiral of green conical sub-units. Each one of these conical sub-units is further comprised of an even smaller spiral of conical sub-sub-units. We could continue this process several times, and eventually we would need a microscope to view the tiny green sub-...-sub-units.

Roughly, the idea of a fractal is that the original object is made up of sub-units, each of which 'look' like the original object, but at a smaller scale. These sub-units are in turn made up of even smaller sub-sub-units, and so on... This phenomenon is what we call self similarity at different scales.

Date: March 1, 2015.
Key words and phrases. Fractals, Iterated Function Systems, Lipschitz Contraction.

Fractal-like objects abound in nature. Examples include:

- Snowflakes
- Coastlines
- Indentations in the earth created by flowing water
- The branches of a tree
- The veins in a leaf
- Cauliflower

Mathematicians are often motivated by phenomenons that occur in nature and they use these phenomenons as inspiration for developing new theories. In this article, we plan to discuss the theory that has been developed for constructing fractal sets.

2. PRELIMINARIES

We first will introduce the notion of a metric. If a and b are two real numbers, we can define the distance between a and b as $d(a, b) = |a - b|$. The absolute value is called a metric, because it allows one to determine the 'distance' between two real numbers. A more abstract metric would be the following. Suppose that $f : [0, 1] \to \mathbb{R}$ and $g : [0, 1] \to \mathbb{R}$ are two continuous functions from the unit interval $[0, 1]$ to the real numbers. We can define a distance, ρ, between f and g by:

$$\rho(f, g) = \max_{x \in [0,1]} |f(x) - g(x)|.$$

With these examples in mind, we now state a formal definition of a metric space.

Definition 2.1. *A metric space (X, d) is a set X and a function $d : X \times X \to [0, \infty)$ such that the following is satisfied:*

- *$d(x, x) = 0$ for all $x \in X$ (the distance between x and itself is zero)*
- *$d(x, y) = 0$ if and only if $x = y$ (the distance between x and y is zero if and only if x equals y)*
- *If $x, y, z \in X$, $d(x, z) \leq d(x, y) + d(y, z)$ (the triangle inequality).*

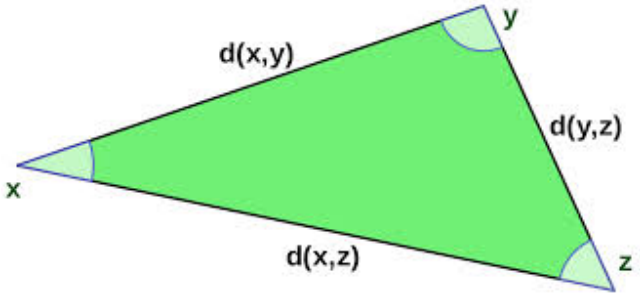


Figure 2: This diagram shows the triangle inequality for the standard metric in $\mathbb{R}^2$.

The first question a mathematician is likely to ask when given a metric space is: 'is it complete'? Perhaps the easiest way to describe completeness is via an example. Suppose we consider the metric space $(0, 1]$ equipped with the absolute value metric. Consider the sequence of points in the metric space $(0, 1]$ listed below:

$$1, \frac{1}{2}, \frac{1}{3}, ..., \frac{1}{n}, ...$$

It is clear that this sequence of points is converging to 0 in the absolute value metric. However, 0 is not an element of $(0, 1]$, and hence we say that $(0, 1]$ is not a complete metric space in the absolute value metric. If we decide to add in the point 0 to our metric space, we get $[0, 1]$, which is a complete metric space. Intuitively, if a sequence of points in a metric space is converging to some point, say x, then x has to be in the metric space if it is complete. We hope that this example and explanation provides an informal intuition for what completeness means. Other examples of complete metric spaces would be:

- $(\mathbb{R}, d)$ where d is the absolute value metric.
- $(C([0, 1]), \rho)$ where $C([0, 1])$ denotes the collection of continuous functions from $[0, 1]$ to $\mathbb{R}$ and ρ is the previously defined metric above.

Suppose now we consider the function $f(x) = \frac{1}{3}x$ on the complete metric space $[0, 1]$. We can observe that if $x, y \in X$,

$$|f(x) - f(y)| = \frac{1}{3}|x - y|.$$

In general, if (X, d) is a metric space, and $L : X \to X$ is some function from X to X with the property that there exists a $0 < C < 1$ such that

$$d(L(x), L(y)) \leq Cd(x, y)$$

for all $x, y \in X$, we say that L is a Lipschitz contraction on X. For the function $f(x) = \frac{1}{3}x$, the value of C is $\frac{1}{3}$. We can think of $f(x)$ as contracting points together because it has a slope strictly less than 1.

An important result in the study of Lipschitz contractions is the well known Contraction Mapping Theorem stated below. We will see that the study of fractals can be thought of as a generalization of this theorem.

Theorem 2.2. *Let (X, d) be a complete metric space, and let $L : X \to X$ be a Lipschitz contraction on X. Then L has a unique fixed point. That is, there exists a unique $x \in X$ such that $L(x) = x$.*

If we return briefly to our example of $f(x) = \frac{1}{3}x$ on the complete metric space $[0, 1]$, we can identify 0 as the unique fixed point of f because $f(0) = \frac{1}{3}(0) = 0$.

We need to discuss three additional concepts before proceeding to the main discussion of the paper:

(1) Compactness: The notion of compactness is a topological property used to describe certain subsets of a topological space (a metric space is the nicest example of a topological space). If we are considering subsets to be contained in the real numbers, compactness of a subset is equivalent to the subset being closed (contains its boundary) and bounded (does not drift off to infinity). We list a few examples:
   - $\mathbb{R}$ is not compact because it is not bounded.
   - $(0, 1]$ is not compact because it is not closed.
   - $[0, 10]$ is compact because it is both closed and bounded.

(2) Measure: Another important notion in the below discussion is that of a measure. Intuitively, a measure assigns a non-negative number to a subset of a space. For

example, the standard (Lebesgue) measure on the real numbers assigns the interval $[a, b]$ the non-negative number $b - a$. Importantly, the notion of a measure is not to be confused with a metric (a measure assigns a non-negative number to a subset, whereas a metric defines a distance between two objects in a space). It just so happens that the distance between a and b is the same as the measure of the closed interval $[a, b]$. A probability measure is a measure which measures the entire space to be 1. For example, if we consider the Gaussian distribution on $\mathbb{R}$ given by

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}},$$

we can define a probability measure, μ, on the subsets of $\mathbb{R}$ in the following way. For a closed interval $[a, b] \subseteq \mathbb{R}$ (or other nice enough subset of $\mathbb{R}$), define

$$\mu([a, b]) = \int_a^b g(x)dx.$$

We note that $\mu(\mathbb{R}) = \int_{\mathbb{R}} g(x)dx = 1$ (this is a famous calculation that is done in many introductory calculus and/or probability classes), and hence μ is a probability measure on $\mathbb{R}$.

3. Constructing the Fractal Set

In 1981, J. Hutchinson published a seminal paper (see [1]), where he generalized the Contraction Mapping Theorem to a finite family, $\mathcal{S} = \{\sigma_0, ..., \sigma_{N-1}\}$, of Lipschitz contractions on a complete metric space (X, d). Indeed, he showed that one can associate to $\mathcal{S}$ a unique compact subset $K \subseteq X$ which is invariant under the $\mathcal{S}$, meaning that

$$K = \bigcup_{i=0} \sigma_i(K),$$

where $\sigma_i(K)$ denotes the image of K under σ_i (or more precisely, the collection of all points $\sigma_i(x)$ for $x \in K$). A finite family of Lipschitz contractions on X is called an iterated function system (IFS) with respect to X, and the compact invariant subset K described above is called the fractal set associated to the IFS. We will briefly describe the different methods that have been discovered for realizing the fractal set. The reader will note that the Contraction Mapping Theorem is used at some point in every method. Therefore, it is appropriate to view the fractal set as a generalization of the Contraction Mapping Theorem.

Before we begin, we would like to present the most standard example of an IFS, and a fractal set. Put $[0, 1]$ as our complete metric space, and consider the two functions $f_0(x) = \frac{1}{3}x$ and $f_1(x) = \frac{1}{3}x + \frac{2}{3}$, which are both Lipschitz contractions from $[0, 1]$ to $[0, 1]$. The IFS is then $\mathcal{S} = \{f_0, f_1\}$.

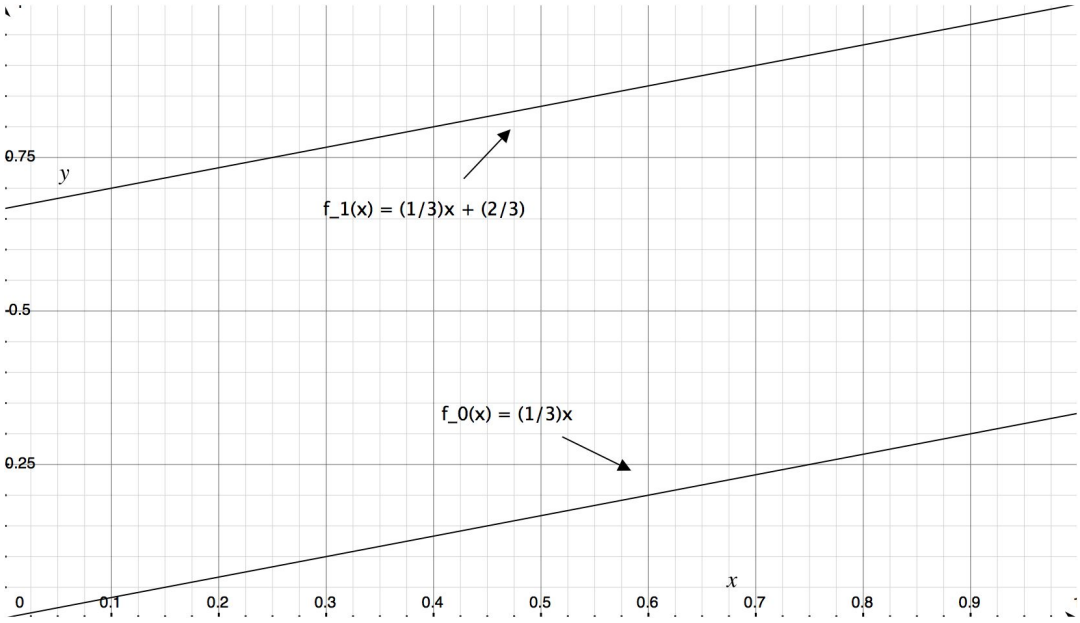


Figure 3: This figure shows the graphs of the two functions in the IFS defined above.

Now, consider the below diagram in which the middle thirds of the intervals are iteratively removed.

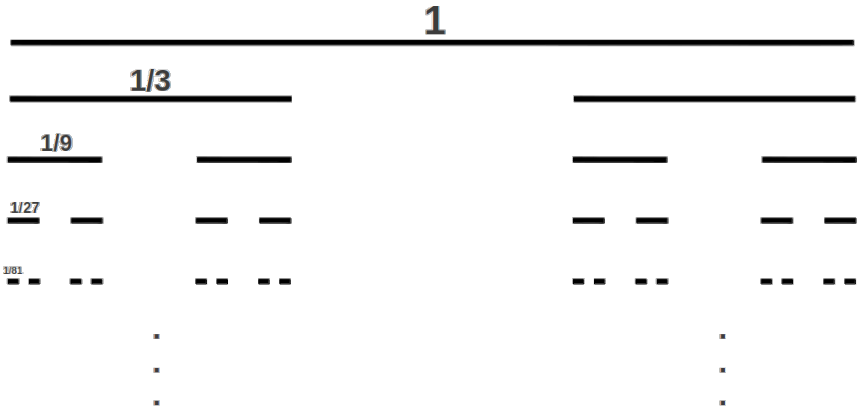


Figure 4: This figure shows the first 4 iterations of removing the middle third.

Imagine now that we repeated this procedure infinitely many times, and we intersected all of these subsets (that is, we considered only the points which belong to every iterate). The reader might wonder if this an empty set (meaning that there are no points left over). However, after some pondering, it is not too hard to see that the endpoints of all of these intervals will belong to every iterate (there will be other points as well). We call this resulting set (the set we obtain after intersecting all of the points) the Cantor set. Let us denote the Cantor set as K. It turns out that K will be the unique compact subset of $[0, 1]$ which satisfies

$$K = f_0(K) \cup f_1(K),$$

Hence, the Cantor set is a fractal set, according to the above description. We can already see from Figure 4 that K will exhibit the self-similarity phenomenon which we discussed in the introduction. For example, the third line in the figure is made up of two sub-units which look like the second line but are smaller by a third.

We will now proceed into the somewhat technical portion of the paper, where we review three ways in which the fractal set can be realized. We take as a starting point an IFS $\mathcal{S} = \{\sigma_0, ..., \sigma_{N-1}\}$ of Lipschitz contractions on a complete metric space (X, d).

(1) The fractal set is the closure of a collection of fixed points:

We first note that since each $\sigma_i : X \to X$ is a Lipschitz contraction, $\sigma_{i_1} \circ ... \circ \sigma_{i_k} : X \to X$ is a Lipschitz contraction as well, where $i_j \in \{0, ..., N-1\}$ for $1 \leq j \leq k$. Therefore $\sigma_{i_1} \circ ... \circ \sigma_{i_k}$ has a unique fixed point in X, which we call $s_{i_1 ... i_k}$. Hutchinson showed in [1] that the fractal set K is the closure of the set of fixed points $s_{i_1 ... i_k}$, for any length tuple $(i_1, ..., i_k)$. This gives a first characterization of the fractal set.

In the Cantor set example above, every endpoint will be a fixed point of $\sigma_{i_1} \circ ... \circ \sigma_{i_k}$ for some tuple $(i_1, ..., i_k)$. Hence, the Cantor set is constructed by taking the closure of the set of all endpoints of the intervals that are iteratively removed.

(2) The fractal set is the fixed subset of the Hutchinson-Barnsley Operator:

If $A, B \subseteq X$, the Hausdorff metric, δ, is defined by

$$\delta(A, B) = \sup\{d(a, B), d(b, A) : a \in A, b \in B\}.$$

We note that $d(a, B) = \inf\{d(a, b) : b \in B\}$ where $\inf$ denotes the infimum (or greatest lower bound). Denote by $\mathcal{K}$ the collection of compact subsets of X. It is known that the metric space $(\mathcal{K}, \delta)$ is complete. The following result is due to J. Hutchinson and M. Barnsley.

Theorem 3.1. [1] [2] *The Hutchinson-Barnsley operator $F : \mathcal{K} \to \mathcal{K}$ given by*

$$K \mapsto \bigcup_{i=0}^{N-1} \sigma_i(K)$$

is a Lipschitz contraction in the δ metric. By the Contraction Mapping Theorem, there exists a unique compact $K \subseteq X$ such that $F(K) = K$. That is,

$$K = \bigcup_{i=0}^{N-1} \sigma_i(K).$$

In other words, what this theorem is saying is that the fractal set K is exactly the unique compact subset of X which is fixed by the Hutchinson-Barnsley operator.

(3) The fractal set is the support of a Borel probability measure:

The fractal set can be realized as the support of a Borel probability measure, μ, on X, called the Hutchinson measure. The support of a measure is the set $\text{spt}(\mu) = X \setminus \{A \subset X : \mu(A) = 0, \text{ and } A \text{ open } \}$. In particular, this measure will be the unique fixed point of a Lipschitz contraction, T, on a complete metric

space of Borel probability measures on X. The metric H is given by

$$H(\mu,\nu) = \sup_{f\in \text{lip}_1(X)} \left\{ \left| \int_X f d\mu - \int_X f d\nu \right| \right\},$$

where $\text{Lip}_1(X) = \{f : X \to \mathbb{R} : |f(x) - f(y)| \leq d(x,y) \text{ for all } x, y \in X\}$, and where μ and ν are Borel probability measures on X. The map T is given by

$$T(\nu) = \sum_{i=0}^{N-1} \frac{1}{N} \nu(\sigma_i^{-1}(\cdot)).$$

In other words, the map T will admit a unique fixed point, which is a measure. This fixed measure μ will satisfy the relation

$$\mu = \sum_{i=0}^{N-1} \frac{1}{N} \mu(\sigma_i^{-1}(\cdot)).$$

and the support of this measure will be K, the fractal set.

This measure theoretic realization of the fractal set has led to further research. Briefly, one can consider the Hilbert space $L^2(K,\mu)$, where K is the fractal set, and μ is the fixed measure. On this Hilbert space, it is possible to obtain functional analytic results about fractals, and iterated function systems, and this is the starting point for my PhD research.

4. ACKNOWLEDGEMENTS

I would like to thank the CU Graduate School for providing a generous grant to the STEMinar. This contribution has allowed the STEMinar to be able to offer a number of grants to its participants. I would also like to thank my advisor, Judith Packer (CU), for her guidance on my research.

REFERENCES

[1] Hutchinson J.,*Fractals and self similarity*, Indiana University Mathematics Journal, **30**, No. 5, 713-747 (1981).

[2] Barnsley M., *Fractals Everywhere* (Second Edition), Academic Press, USA,1993.

E-mail address: Trubee.Davison@colorado.edu

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF COLORADO, BOULDER, COLORADO

EXPLORING STRONGLY INTERACTING SYSTEMS THROUGH THE GAUGE-GRAVITY DUALITY

OSCAR HENRIKSSON

ABSTRACT. I give a short introduction to gauge-gravity duality, an active research area in theoretical high-energy physics. Some background is given on string theory, through the study of which gauge-gravity duality was discovered. The key ideas of this duality are presented and some of the applications are discussed. Finally, I touch upon the particulars of my own research. I avoid technicalities to a great extent, opting for a more conceptual approach.

1. BACKGROUND: STRING THEORY AND UNIFICATION

Theoretical physics has long been occupied with constructing theories unifying diverse natural phenomena. Newton's theory of gravity described both the planets and a falling apple within the same framework, Maxwell's equations unify the magnetic and electric forces, and quantum field theory unifies Einstein's special theory of relativity and quantum mechanics; the list can be made much longer. Currently physicists believe the fundamental building blocks of nature consists of a handful of elementary particles, including among others electrons and quarks, and four fundamental forces or interactions. String theory is a continuation in the quest of unification: it attempts to describe all of these known elementary particles and fundamental forces within a single theoretical framework.

The details and workings of string theory are not the focus of this paper. It is sufficient to say that string theory provides a rich and complex mathematical framework that predicts many interesting physical phenomena and does contain a quantum mechanical theory able to describe all of the fundamental particles and interactions, but that it hasn't yet been experimentally verified. However, even if string theory does not end up being a theory that describes nature, the study of it has been fruitful and has led to many new insights and tools in theoretical physics. It is one of these "applications" of string theory that is to be our focus here, namely gauge-gravity duality.

2. GAUGE-GRAVITY DUALITY

As mentioned, one of the foremost objectives of theoretical physics today is to unify the four known fundamental interactions in one consistent framework; of these four interactions, gravity is the one that has proven the toughest nut to crack. A fully consistent theory that incorporates both gravity and quantum mechanics is hard to come by. String theory, however, is such a theory, which is one reason why it has become so popular among theoretical physicists. It was when studying some specific versions and limits

Date: January 25, 2015.

Key words and phrases. Theoretical physics, high-energy physics, gauge-gravity duality, AdS/CFT correspondence.

of string theory that researchers stumbled upon what became known as gauge-gravity duality. This surprising discovery provides a map between the two types of theories that string theory incorporates, namely quantum field theories (QFTs) and gravitational field theories. The latter are similar to Einstein's theory of gravity, consisting of a classical field theory describing a curved spacetime. QFT, on the other hand, is a quantum mechanical framework used to describe systems with many degrees of freedom. It is used heavily in various fields of physics, including elementary particle physics, nuclear physics, and condensed matter physics. Most QFTs are complicated theories, and few of them admit exact solutions. The usual tools of the trade are perturbative approximation techniques that only work when interactions are sufficiently small. This, however, is not the case for many important systems, creating a need for different approaches.

Gauge-gravity duality was first conjectured to exist by Maldacena [1], with details further developed in papers by Witten [2] and Gubser *et al.* [3]. The discovery of this correspondence generated a flurry of excitement that continues to this day; the original paper has over 10.000 citations to date. The duality is also known as the AdS/CFT correspondence, a name that is derived from the two theories involved: The gravitational theory is set in anti-de Sitter (AdS) space, which is a solution to Einstein's field equations with negative spacetime curvature. The QFTs in the correspondence are of a specific type known as conformal field theories (CFTs); theories of this type do not have a preferred length scale. Most uses of the correspondence involve doing calculations in the gravitational theory and translating the results into information about the QFT, thus providing knowledge about a strongly interacting quantum system which would otherwise be very difficult to acquire. Interestingly, the gravitational theory takes place in one spacetime dimension higher than the QFT; if, for example, we want to study a QFT in three dimensions, we must do calculations in a gravity theory in four dimensions. Because of this strange property, methods involving gauge-gravity duality techniques sometimes go by the name of "holography".

3. Applications of the duality

The duality itself is a fascinating piece of theoretical physics that is worth studying on its own, but it also has possible applications that could prove very useful. Most of these applications involve understanding the properties of strongly interacting systems. This is one of the most urgent problems in theoretical physics today. Related questions are pondered in such diverse subfields as condensed matter physics, nuclear physics, and elementary particle physics. Possible solutions could have applications to a wide range of experimental observations such as the quark-gluon plasma, strange metals, and high-temperature superconductivity. However, this is a very challenging problem; strong interactions often mean that traditional theoretical tools like perturbative expansions and quasi-particle formulations do not work, and exact solutions can be obtained only in a few special cases. Thus, even though the solutions to these puzzles have been sought for decades and have potentially far-reaching implications, there are still many open problems.

Crucially, in gauge-gravity duality, if one of the theories in the duality is strongly interacting, the other is instead weakly interacting and thus easier to work with. Hence, calculations can be preformed in the weakly interacting theory and the results can then

be translated back to the strongly interacting theory through the duality map. For the applications that will be mentioned here this is taken advantage of by doing calculations on the gravity side of the duality and then mapping the results over to the QFT side. This can provide information about much that is hard to access with traditional methods, such as real-time and non-zero density behavior.

A first important example of the duality's applications is heavy ion collisions [4]. The QFT in question here is quantum chromodynamics, the theory of the strong nuclear force, which as its name implies is indeed strongly interacting in certain energy regimes. Currently active experiments, such as those preformed at the Large Hadron Collider at CERN in Switzerland, are probing some of these regimes, and a theoretical description is therefore important.

My research focuses on applications to another field, condensed matter physics [5], where the ease of studying systems at non-zero density is an important advantage of duality methods. Current experiments in this area are studying phenomena such as non-Fermi liquids, topological insulators, and high-temperature superconductors, and a complete theoretical description is lacking. Using the duality, models of superconductors have been constructed [6]. Furthermore, properties of fermions have been studied [7] and certain theories have been found to display non-Fermi liquid behavior; more on this below.

My work, in collaboration mainly with my advisor Prof. Oliver DeWolfe, focuses on so-called "top-down", or exact, examples of gauge-gravity duality. Here, top-down means that the correspondences we work with can be derived directly from string theory. Thus we know exactly what the two theories in the correspondence are, allowing us to make powerful statements about the details of the microscopic physics. This is to be contrasted with bottom-up models, which postulate a gravity theory and work out the consequences, without knowing exactly which QFT the results describe. Bottom-up models are important since they are simpler to work with and provide good phenomenological models, but they are on looser theoretical footing than their top-down counterparts.

More specifically, we are currently studying fermions in a specific QFT known as ABJM theory. This theory is set in two spatial dimensions and is very special; in a certain sense it is the most symmetric QFT in that number of dimensions. And while it is true that most QFTs describing real-world systems do not have this degree of symmetry, it makes this theory a lot easier to deal with while retaining many properties of more realistic theories. Furthermore, a theory with two spatial dimensions is interesting for condensed matter applications, since many compounds studied in this area are effectively two-dimensional, or flat.

We study the fermions in the ABJM theory at non-zero charge density, which is something that is hard to do using other leading approaches to strongly coupled QFTs such as computer simulations. Fermions are subject to the Pauli exclusion principle; at non-zero density this leads them to form Fermi surfaces, and the properties of these surfaces and nearby excitations explain many features of real-world materials. In the standard theory of metals (Fermi liquid theory), the excitations near the Fermi surface are stable, meaning they can be described by standard perturbative methods. However, in many materials of current interest in physics such as the strange metals observed in

high-temperature superconductors, the excitations are not stable and the perturbative description breaks down. This non-Fermi liquid behavior is no obstacle for duality methods, however, and they can therefore provide good theoretical descriptions. Indeed, in my advisor's previous work it has been shown that gauge-gravity duality predicts non-Fermi liquid behavior in the super-Yang-Mills theory in three spatial dimensions [8], and we obtained similar results in our first comprehensive study of the two-dimensional ABJM theory [9]. This is but one of several examples of how these surprising methods coming from high-energy physics might be able to describe condensed matter systems.

4. Further reading and acknowledgements

This paper has been an all too brief introduction to gauge-gravity duality and some of its applications. Many more resources are available for the curious reader. For a non-technical introduction, see for example [10]. For the working theoretical physicists, [11] provides a comprehensive review.

I would like to thank Trubee Davison for promoting connections across different STEM fields by arranging many interesting STEMinars at the University of Colorado Boulder.

References

[1] Juan Martin Maldacena, *"The Large N limit of superconformal field theories and supergravity"*, Adv.Theor.Math.Phys. 2 (1998) 231-252.

[2] Edward Witten, *"Anti-de Sitter space and holography"*, Adv.Theor.Math.Phys. 2 (1998) 253-291.

[3] S.S. Gubser, Igor R. Klebanov and Alexander M. Polyakov, *"Gauge theory correlators from noncritical string theory"*, Phys.Lett. B428 (1998) 105-114.

[4] Oliver DeWolfe, Steven S. Gubser, Christopher Rosen and Derek Teaney, *"Heavy ions and string theory"*, Prog.Part.Nucl.Phys. 75 (2014) 86-132.

[5] Sean A. Hartnoll, *"Lectures on holographic methods for condensed matter physics"*, **arXiv:0903.3246 [hep-th]** (2009).

[6] Sean A. Hartnoll, Christopher P. Herzog and Gary T. Horowitz, *"Holographic Superconductors"*, JHEP 0812 (2008) 015.

[7] Thomas Faulkner, Hong Liu, John McGreevy and David Vegh, *"Emergent quantum criticality, Fermi surfaces, and AdS(2)"*, Phys.Rev. D83 (2011) 125002.

[8] Oliver DeWolfe, Steven S. Gubser and Christopher Rosen, *"Fermi surfaces in $\mathcal{N} = 4$ Super-Yang-Mills theory"*, Phys.Rev. D86 (2012) 106002 or **arXiv:1207.3352v2 [hep-th]** (2012).

[9] Oliver DeWolfe, Oscar Henriksson and Christopher Rosen, *"Fermi surface behavior in the ABJM M2-brane theory"*, **arXiv:1410.6986 [hep-th]** (2014).

[10] Juan Martin Maldacena, *"The Illusion of Gravity"*, Scientific American November 2005.

[11] O. Aharony, S.S. Gubser, J. Maldacena, H. Ooguri and Y. Oz, *"Large N Field Theories, String Theory and Gravity"*, **arXiv:hep-th/9905111v3** (1999).

E-mail address: Oscar.Henriksson@Colorado.edu

Department of Physics, University of Colorado, Boulder, Colorado

'EMBODIED' REPRESENTATIONS OF EMOTION IN THE BRAIN

MARIANNE REDDAN

ABSTRACT. Colloquially, 'emotion' and 'feelings' share semantic meaning, but it remains unknown whether the expression of emotion and bodily sensation share representational space in the human brain. Do emotional experiences include activation in sensory processing regions specialized to represent the human body? There is growing evidence that emotions are 'embodied,' or grounded in some modality, such as perception (e.g., of the five senses) and action (e.g., movement, proprioception). However, a causal link between self-reports of bodily sensations and discrete emotional states has not been established. This proposal investigates whether a person's subjective emotional state can be predicted by neural activation patterns in cortical regions that encode body-specific information: the somatosensory, motor, and insular cortices. Sensitive and specific signatures of emotion will increase our understanding of how emotion influences behavior, and may provide powerful diagnostic tools for emotion disorders in the clinic.

1. INTRODUCTION

Emotions are an essential part of the human experience: They color our daily lives with rich sensations, contain content, and contribute to motivation, goal selection and attainment [1]. Despite the common notion that emotions are felt, computational models of emotion historically treat emotional concepts as abstract symbols, encoded independent of sensory modalities [2,3]. However, evidence for these models have not been supported by neuroimaging: Emotion-related neural activation is neither amodal nor localized [4,5]. Indeed, self-generated feelings of sadness, anger, happiness, and fear elicit distributed activation patterns across the whole brain, including the brainstem [6]. If emotion processing is not reducible to a single section of cortex, what are the critical components of emotion representation in the brain?

1.1. **Embodied Emotion Theory.**

Recent work proposes that emotions emerge from highly distributed multimodal neural representations that include, but are not limited to, somatic perception, cognitive appraisal (or psychological description and evaluation), and working memory [7]. These theorists contest that emotions are not abstract symbols, separated from sensory processing, but are instead highly grounded in bodily sensations, perceptual simulations, and contextual information [8]. For example, a negative experience occurs: A reckless driver cuts you off during rush hour traffic. This makes you angry. The brain encodes this experience through many perceptual modalities and stores an integrated multimodal representation of this experience in memory (e.g., the color of the car and the sounds of traffic, the action of shaking your fist at the perpetrator, internal reflections on threat and anger). Later, when you encounter

Date: January 31, 2015.
Key words and phrases. Psychology, neuroscience, emotion, embodiment, fMRI, machine learning.

another speeding red car, it is hypothesized that these multimodal representations are simulated, or reenacted, in the perceptual, motor, and introspective circuits recruited during the initial experience [9]. Furthermore, the experience of anger in other situations might manifest as tension in the muscles of the arms and fist, and an increase in heart-rate and action-readiness [10]. In this way, the experience of 'anger' cannot be isolated from experiences with the world and internal states; instead, anger emerges from dynamic and distributed representations activated throughout the brain and body [1].

Niedenthal (2008) proposes that there is a "reciprocal relationship between the bodily expression of emotion and the way in which emotional information is attended to and interpreted [11]." This framework for the embodiment of emotion arises from observations that emotional states can be triggered by physiological feedback, and that bodily sensations are consistently reported during emotional experiences. For example, the restriction of motor movements, such as preventing facial expressions by asking participants to hold a pen in between their teeth, interferes with interpretations of emotional images and the participant's subjective experience of emotion [12,13]. Indeed, somatic information like heart-rate and posture can influence the emotional interpretation of a new situation: Participants sitting in an upright position report greater feelings of pride when they are delivered good news than their slumped over counterparts [14].

In embodiment theory, the assertion is that the neural representation of the bodily state is an intrinsic component to the emotional experience. Seeking support for this reciprocity, Nummenmaa, Glerean, Hari, and Hietanen (2013) provided subjects with emotional words, stories, movies, and facial expressions and asked their participants to color a map of the human body, in order to indicate where they felt increases or decreases in sensation. Using classification algorithms, they were able to produce independent topographical bodily maps of emotions. Furthermore, these predictive maps of emotion sensation were accurate across both Eastern and Western cultures, suggesting some universality to emotion generation and experience [15]. While this is a striking example in support of emotion embodiment, it does not provide a causal link between brain and behavior. Furthermore, participants in this study were required to report bodily sensations; if participants had the choice to report the lack of bodily sensation, would the results have been different? Because these topographic maps were derived from self-report data, we cannot be sure if emotion-predicting body parts are actually included in the neural representation of emotion or if they are a by-product of rich associations between the human body and emotion in language.

1.2. **Scientific Aims.** This proposal aims to investigate the causal link between an emotional experience and bodily sensations by analyzing activity patterns in neural regions known to be responsible for three components of bodily sensation: representation of individual body parts in the somatosensory cortex, motor planning (premotor and motor cortices), and visceral representations of inner organs (insular cortex). Can activation in these regions reliably predict an emotional state? With the advent of powerful pattern classification techniques and high resolution multiband imaging there is a new opportunity to investigate the embodiment of emotions with increased sensitivity and specificity. Through a visual emotion induction paradigm, and the collection of neural activation and self-report data, this investigation will address the follow questions:

- Are the findings of Nummenmaa, Glerean, Hari, and Hietanen (2013) robust in our sample? Will participants report the same patterns of bodily emotion representation? For example, will participants identify sensation in their fists when they feel anger?
- Can we reliably predict self-reported bodily sensations from neural activation patterns in regions specific to the representation of the human body? For example, will activity to the somatosensory cortex predict self-reported sensations of one's fist?
- More specifically, will these predictive neural signatures trained on self-reported bodily sensation reflect established somatotopy? For example, the somatosensory cortex contains a neural map representing every body part. Will activation in the neural space specific for hand representation correlate with self-reports of fist sensation during an experience of anger?

The answers will provide insight to the representation of emotion, and broaden our understandings of brain-body interactions. Furthering our knowledge of the neural basis of emotion will impact both health and welfare, as it may inform the diagnosis and treatment of emotional disorders.

2. DESIGN

2.1. **Method.** Participants are shown a series of 112 highly arousing emotional images that vary in valence, level of scene complexity, and situational content, while lying inside of a functional magnetic resonance imaging (fMRI) scanner. Picture stimuli were selected from the International Affective Picture System System [16] and the Geneva Affective Picture Database [17] using published normative arousal and appraisal ratings [18,19] so that each image was selected to be of high arousal and belonged to a predetermined category of appraisal (see Appendix A, Figure 1). Image presentation lasts 4-seconds, and are spaced in time by a randomized jittered inter-trial-interval of 2 to 12-seconds. The image presentations are divided into two 7.5-minute runs, where the first 56 images are presented in a randomized order, followed by a random presentation of the remaining 56 images. The MATLAB Psychtoolbox package is used for the stimulus presentation. Pupil dilation, skin conductance, and pulse rate are recorded throughout the task. Scans are acquired using a high temporal and spatial resolution multiband imaging protocol with a 460-ms repetition time (TR) in a Siemens 3-Tesla magnet.

In order to control for motor-response tendencies and meta-awareness of emotion, self-reports are collected after the imaging session. Participants are shown the images again, in the same order presented to them during the experiment, outside of the scanner, and are asked to color on a body map, where, if at all, they feel emotion in response to the image in their body. Subjective interpretations of the emotional content and personal experience of the viewer (i.e., appraisals) are also collected outside of the scanner. Each image is rated on the seven predetermined appraisal dimensions (see Table 1). Ratings are collected on a visual analogue scale with the anchors (NOT AT ALL and EXTREMELY SO). Participants can mark anywhere on the scale and submit their rating with a click. Ratings are encoded using a 1 - 100 scale.

2.2. **Analysis Plan.** The first step is to replicate the findings of Nummenmaa, Glerean, Hari, and Hietanen (2013) in our self-report data. Utilizing subject-specific emotion appraisal ratings, we will classify the body maps, in order to yield maps of body sensation unique to emotional content.

The self-reported body maps will be divided into six sections: arms, legs, groin, lower abdomen, chest, and head and neck. For each subject, per stimulus image, the maximally represented body part will be extracted and input as a regressor into the model for imaging analysis, so that there will be six regressors predicting unique bodily sensation in a first level general linear model. Using machine learning classification techniques, trained on the resulting beta images within only three predefined neural regions of interest (somatosensory, motor, and insular cortices) from the first functional run, and their corresponding body category, six predictive weight maps for emotion-related bodily sensation will be developed. The signatures will be cross-validated using a leave-one-subject-out procedure. These six signatures will then be tested on each trial of a hold-out set comprised of each subject's second functional run (the second set of 56 images). The signature with the highest pattern expression, calculated as the dot product of the weight map and the test image, will be used to test the accuracy of the classifiers (where chance level will be set at 100/6 or 16.7%). This analysis will reveal if activation in body-representation regions during an emotional experience can predict self-reported bodily sensations. If it does, a causal link between neural activation during emotion and bodily sensation will be established.

The final analysis aims to 'decode' neural activation in bodily-representation regions with finer-grained detail. The primary motor cortex and somatosensory cortex are somatotopically organized to represent contralateral body parts. This organization holds a high degree of intro- and interindividual reproducibility [20]. Using an anatomical mask of somatotopic organization to define the six body categories (arm, leg, groin, lower abdomen, chest, and head and neck), we can extract beta weights indicating the magnitude of activation within each subregion within individual subjects. We can then correlate activation magnitude within body part category to the magnitude of self-reported sensation in that region. This analysis will provide further support that discrete somatosensory representations are shared with representations of emotional experiences, such that experiencing anger does indeed involve activation in the neurons responsible to balling your hand into a fist. Subsequent analyses may be performed in which we attempt to predict a self-reported emotional experience from neural activation in these body-specific sensory cortices.

3. IMPLICATIONS

Human emotion is a complex phenomenological experience that emerges from neurobiological and psychological processes. Despite its research popularity and societal importance, emotion remains elusive to neuroscience: Neither the neural structure of an emotional experience, nor the processes that yield it are well understood [21,22,23]. However, the question of how emotions are represented in the brain is urgent: Psychological disorders rooted in emotional disturbance, such as mood, anxiety, personality, and substance abuse disorders, plague over 42.2% of Americans and cost the United States over $42 billion a year in federal spending [24]. These disorders are as serious

and debilitating for society and the economy as they are for the individual. Furthermore, current means for both treatment and diagnosis are insubstantial because they lack both diagnostic accuracy and reliability [25]. The development of standardized measures of emotion, such as neural and physiological biomarkers, that can be shared and compared across studies holds promise for the clinical application of emotion research. This investigation will aid in the biomarker initiative through the development of predictive neural signatures of emotion expression which will be unique to how discrete emotions are 'felt' in the body. Previous attempts to classify emotion have accurately captured emotion-related high and low arousal, and positive and negative valence [4,5], but distinctions between high and low arousal states, and bipolar valence states, are not enough to explain a complete, content-rich emotional experience. Advancements in the classification of neural expressions of emotion are critical to our understanding of the neurobiological mechanisms that create and modify the emotional experiences serving both human success and dysfunction. Furthermore, this research can inform psychiatric diagnosis of emotion-related disorders, and inspire a new avenue of treatment targeting bodily sensations.

4. Acknowledgements

This project is made possible by the support and guidance of my advisor, Tor Wager, PhD, Luke Chang, PhD, and the members of the Cognitive Affective Neuroscience Laboratory in the Department of Psychology and Neuroscience at the University of Colorado-Boulder. McKell Cartson, PhD, and June Gruber, PhD, have also provided critical feedback and support.

Appendix A.

Figure 1.

Table 1.

Appraisal Rating Dimensions

a.	level of contamination or disease
b.	degree of physical threat present
c.	level of intentionality of scene
d.	level of plesant or unpleasantness
e.	degree of moral or immoralism of scene
f.	degree of empathy felt for actors in scene
g.	degree to which the scene creates a narrative

B)

REFERENCES

[1] Barrett LF, Mesquita B, Ochsner KN, Gross JJ (2007) The Experience of Emotion. Annual Reviews in Psychology(58)373-403.
[2] Fodor J. (1975) The Language of Thought. Harvard University Press. Cambridge, MA. Newell, A (1980) Physical Symbol Systems. Cognitive Science 4(2):135-183.
[3] Chang, LJ., Gianaros, PJ, Manuck, SB, Krishnan, A, Wager TD (submitted) A sensitive and specific neural signature for picture-induced negative affect.
[4] Chikazoe J, Lee DH, Kriegeskorte N, Anderson AK (2014) Population coding of affect across stimuli, modalities and individuals. Nature Neuroscience(17)8:1114-1125.
[5] Damasio AR, Grabowski AB, Damasio H, Ponto LLB, Parvizi J, Hichwa RD (2000) Subcortical and cortical brain activity during the feeling of self-generated emotions. Nature Neuroscience 3, 1049-1056.
[6] Thagard P and Aubie B (2008) Emotional consciousness: A neural model of how cognitive appraisal and somatic perception interact to produce qualitative experience. Consciousness and Cognition 17811-834.
[7] Wilson-Mendenhall, CD, Barrett LF, Simons, KW, and Barsalou LW (2011) Grounding emotion in situated conceptualization. Neuropsychologia 49:1105-1127.
[8] Barsalou LW (2008) Grounded Cognition. Annual Reviews in Psychology. 59:617-45.
[9] Niedenthal PM (2007) Embodying Emotion Science 316, 1002.
[10] Niedenthal, PM, Barsalou, LW, Ric, F, Krauth-Gruber S (2005) in Emotion: Conscious and Unconscious, L. F. Barrett, P. M. Niedenthal, P. Winkielman, Eds. (Guilford, New York), pp. 21Ð50.
[11] Effron, DA., Niedenthal, PM, Gil, S, Droit-Volet, S (2006) Embodied temporal perception of emotion. Emotion 6(1) 1-9.
[12] Strack F, Martin LL, and Stepper S (1988) Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. Journal of Personality and Social Psychology, 54(5),768-777.
[13] Nygaard, L.C. and Lunders, E.R.(2002). Resolution of lexical ambiguity by Emotional tone of voice. Memory and cognition, 30, 583-93.
[14] Nummenmaa, Glerean, Hair, and Hietanen (2013) Bodily maps of emotions. PNAS (111) 2:646-651.
[15] Lang, P.J., Bradley, M.M., and Cuthbert, B.N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8. University of Florida, Gainesville, FL.
[16] Dan-Glauser ES and Scherer KR (2011) The Genevea affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance. Behavioral Research 43:468-477.
[17] Mikels JA, Fredrickson BL, Larkin GR, Lindberg CM, Maglio SJ, Reuter-Lorenz PA (2005) Emotional category data on images from the International Affective Picture System. Beahvior Research Methods 37(4), 626-630.
[18] Libkuman TM, Otane H, Kern R, Viger SG, Novak N (2007) Multidimensional normative ratings for the International Affective Picture System. Behavior Research Methods. 39(2), 326-334.
[19] Alkadhi H, Crelier GR, Boendermaker SH, Golay X, Hepp-Reymond MC, and Kollias SS (2002) Reproducibility of Primary Motor Cortex Somatotopy Under Controlled Conditions. American Journal of Neuroradiology 23:1524-1532.
[20] Barrett LF (2006) Are Emotions Natural Kinds? Perspectives on Psychological Science (1) 1:28-58.
[21] Izard CE (2007) Basic emotions, natural kinds, emotion schemas and a new paradigm. Perspect Psychol Sci 2: 260Ð280.
[22] Panksepp J (2007) Neurologizing the psychology of affects: How appraisal-based constructivism and basic emotion theory can coexist. Perspect Psychol Sci 2(3): 281Ð296.
[23] Anxiety and Depression Association of America (1999) Report. http://www.adaa.org/about-adaa/press-room/facts-statistics.

[24] Kapur S, Phillips AG, and Insel TR (2012) Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? Molecular Psychiatry(12):117-9.

E-mail address: Marianne.Reddan@colorado.edu

DEPARTMENT OF PSYCHOLOGY AND NEUROSCIENCE, UNIVERSITY OF COLORADO, BOULDER, COLORADO

A NEW MANUFACTURING PARADIGM FOR ADVANCED METAL - CERAMIC COMPOSITES

FNU SHIKHAR

ABSTRACT. Cermets (composite of ceramics and metals) have shown unprecedented properties of ceramics (with standing high temperature and hardness) and metals (toughness and thermal conductivity to avoid thermal shock) but most of the ceramics requires high temperature to sinter (to form solid mass from compacted powder) at which metals (except platimum) either melts or oxidizes. This has been the crucial reason why Cermets have been proven tough to manufacture. In recent time, a new process has been developed to sinter ceramics where electric field is applied while they are being heated up. This process has been named, flash sintering, which has reduced the temperature needed to sinter by hundreds of degree centigrade and the process in itself is very fast. Further understanding has shown that electric field alters the whole mechanism of sintering by creating defect pairs of interstitials and vacancies which is responsible for this phenomena. In this work we have tried to establish, that under the effect of field the metal sticks well to the ceramic (wetting) which can lead to synthesis of metal-ceramic composites.

1. INTRODUCTION AND BACKGROUND

Ceramics have a wide range of electronic, optical, structural and chemical properties. Having high strength and being chemically inert they are used in bio-mechanical applications such as bone replacement (alumina). At high temperatures ceramics are employed as thermal barrier coatings on turbine blades (zirconia). In energy applications they are used in solar cells (titania).

Because ceramics have high melting points they cannot be cast, forged and machined like metals. Instead they are made by the sintering process. In this process fine powders of the ceramic material are pressed into shape, and then heat treated at very high temperatures. Mass transport by solid state diffusion leads to densification over a period of several hours at temperatures that are typically in the 1300 degrees Celsius to 1600 degrees Celsius range.

It has been discovered recently that by applying modest electrical fields ceramics are sintering in mere seconds at temperatures that are several hundred degrees below conventional sintering. For example 3 mol % yittria stabilized zirconia (3YSZ), which required several hours at 1450 degrees Celsius in conventional sintering, reaches full density in just 2 Ð 3 seconds at a furnace temperature of only 850 degrees Celsius [1]. This new process has become known as 'flash sintering'. Thus, the energy consumption is reduced by more than 90% with respect to conventional sintering, as shown in figure 1. Recent scientific investigations suggest that flash sintering involves the generation and mobility of atomic defects, which enables ultrafast sintering at low temperatures

Date: January 31, 2015.

[2]. Therefore flash sintering is emerging as a new paradigm in the manufacturing of ceramics.

The phenomenon of flash sintering is demonstrating other unexpected advantages. Multiscale, multilayered, mulitphase architectures, which normally cannot be co-sintered by conventional methods because different phase sinter at different rates, also known as constrained sintering, have been shown to be processed easily by the flash method [3,4].

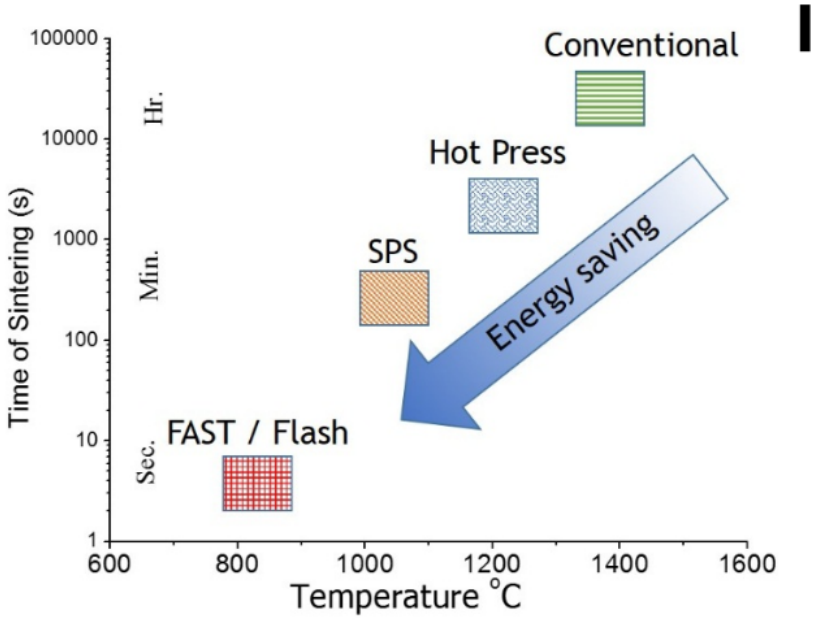


Figure 1. Trend of saving energy in terms of temperature and time with new methods. Flash sintering is the most efficient one.

The present work shows the effect of application of flash sintering to Cermets. They possess the best properties of both ceramics and metals; namely high temperature stability and stiffness of ceramics, and ductility, toughness and higher thermal conductivity of metal. Their applications are found in thermal barrier coating to withhold higher temperatures (because of their higher thermal conduction), as joining part of metal and ceramics or sealants for their ductile nature, as precision cutting tools and in bio-medicals for bone replacement. However, sintering ceramics and metal powder together is not easy since metals melt and/or oxidize at temperatures much below that needed for ceramic particles to sinter. This is where flash sintering comes to aid: to sinter the composite at low temperature without oxidizing the metallic part.

2. Experiments

We take the system of 3 mol % yittria stabilized zirconia (3YSZ) with 15 vol% of aluminum metal added into it. 3YSZ powders have nanosize particles of dimension 70nm while aluminum metal particles are of the order of 1-2 μm. Slurry of 3YSZ and aluminum in water with some binders are made in order to mix them together and ultrasonicate it to homogenise the distribution of metal in ceramics. Then after the solution

kept in the drying over at 80 degrees Celsius to get the homogenous powder back and crushed back into fine powder to be compressed in a uni-axial compression die to give it a shape of dogbone.

The setup of the experiment has been explained in the work of Jha and Raj [1,4], figure2. Dogbone shaped specimen (with initial density 50% of the theoretical value) were lowered into a furnace with the help of platinum wire which also act as the electrodes to apply electric field. These electrodes are connect with the power supply to apply voltage or current. Figure 3 shows the response of applying an electric field of 175V/cm on the specimen at the furnace temperature of 830 degrees Celsius. At this temperature, under the effect of electric field the current starts to rise and continues to rise until it reaches the final current limit set by the power supply, since one can only control current or voltage across a resistance so the power supply switches to current control mode. The voltage drops to a value determined by the conductivity of the specimen and current limit, according to the Ohm's law. The sample shrinks quickly and reaches near 90% of its theoretical density.

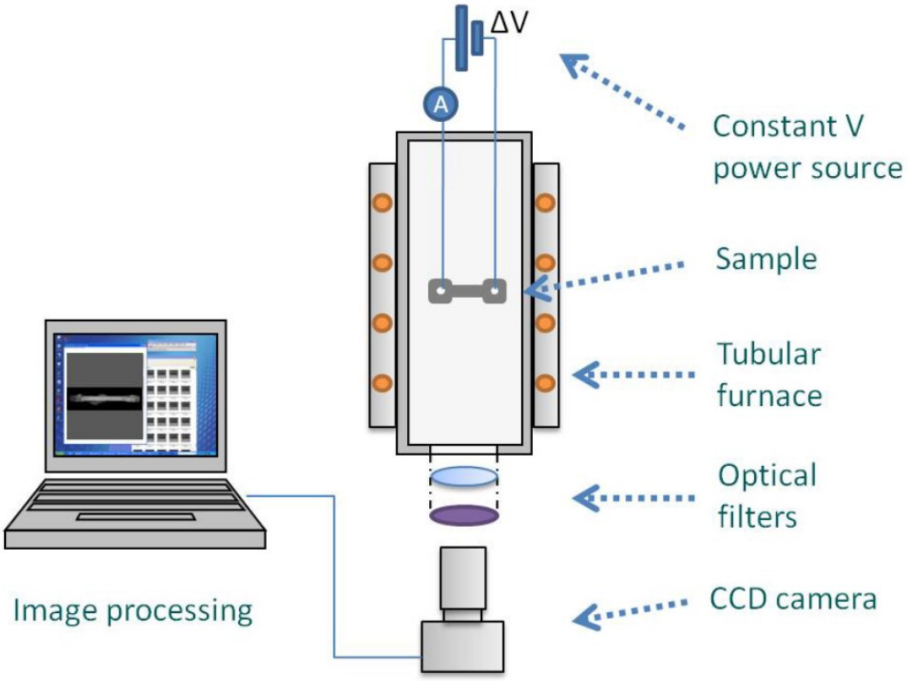


Figure 2 shows the setup of the experiment where a sample is kept in the furnace. The shrinkage is measured from bottom using the camera and the voltage is applied using a power source. A computer is used to record the voltage, current, temperature and shrinkage in the specimen.

The sintered specimen is then cooled down, and cut across the cross-section to be seen under SEM (scanning electron microscope), as shown in figure 4. The sample preparation for SEM involves polishing the surface up to 1μm finish. The thermal etching of sample was avoided because that would lead to oxidation of metal part. The sample is then sputtered with gold-paladium to the thickness of 3 nm. Other part of

the specimen was crushed into powder and scanned under the x ray diffraction to be compared with green powder. The scan was done from 2 theta 20 to 50, figure 5.

3. Results and Discussion

The experimentation plot, figure 3, shows transition from voltage control to current control with an incubation period of 25 sec, under the electric field of 175 V/cm. Current limit of 75mAmm^{-2} is used in order to get full densification. An anomaly is seen when compared to usual ceramics. Right after the transition from voltage controlled to current controlled, the voltage drops to a certain value and then comes up a little and continues to be same for rest of the part of flash sintering that lasted for about 30 sec. It can be probably explained in terms of oxidation of aluminum metal into alumina, since alumina is an adherent layer so it stops the further oxidation and alumina is insulator.

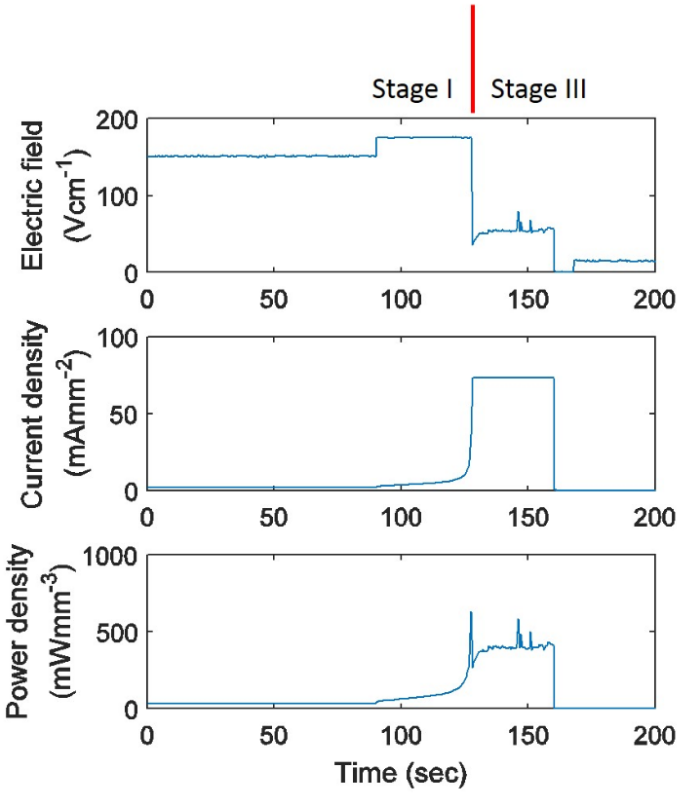


Figure 3. The electric response of specimen under flash condition. The furnace temperature of the experiment was kept at 830C for all the experiments.

On looking at the micrograph, figure 4, we see no pores at the interfaces of metal and ceramics. EDS analysis confirm the two phases being Al metal and zirconia ceramics. An enlarge picture around one of the aluminum particle shows the nano-scale penetration of metal into ceramics.

Xray diffraction confirms that although the sintering is happening at a much higher temperature than melting point of aluminum, but about half of the aluminum is still preserved in metal form. We have done some hardness testing, under vicker load of 1000g held for 15 sec on pure 3YSZ and metal-3YSZ cermet and they seem to have a

comparable values: 13 ± 0.74 GPa and 11.66 ± 0.60 GPa. Since none of the sample showed any cracks at the tip of the indentation so toughness comparison was not possible, but it is sufficient to state that the material have comparable hardness (considering same experiment showed the hardness of iron being 300 GPa).

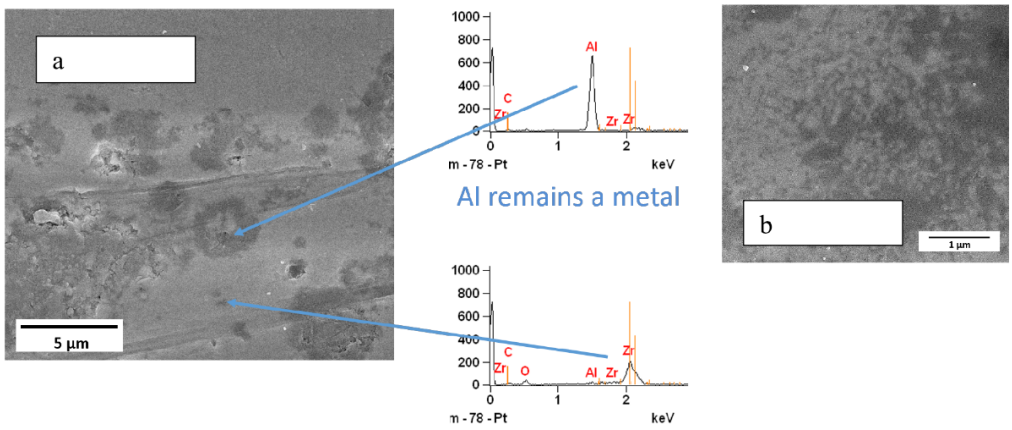


Figure 4: Co-sintering of metal and ceramics, composite, under the flash condition of 175V/cm and 75mAmm^{-2} kept for 30sec.

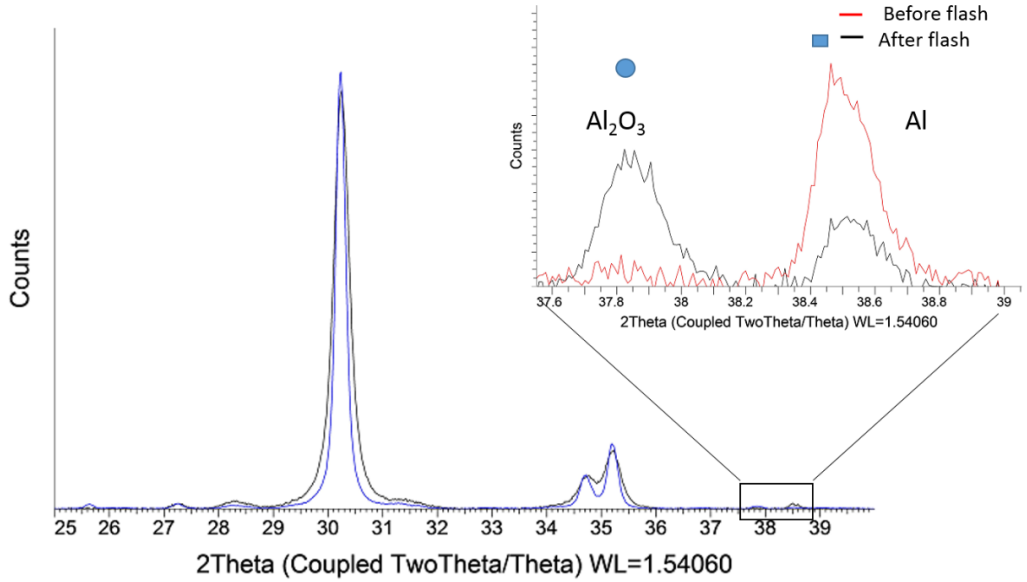


Figure 5: XRD pattern of metal ceramic composite. 2theta 37.5 to 39 is enlarged to show the residual Aluminum metal in the cermet before and after the flash sintering.

4. Conclusion

- It is possible to sinter ceramic-metal composites under the electric field.
- The ceramics shows hardness comparable to that of pure 3YSZ.
- Microstructure nanoscale inter-penetration of metal into ceramics which is the evidence for good wetting of metal with ceramic surface.

REFERENCES

[1] Cologna, M., Rashkova, B. and Raj, R. Flash Sintering of Nanograin Zirconia in <5 s at 850ąC. J. Am. Ceram. Soc. 93, 3556Đ3559 (2010).
[2] Raj, R. Joule heating during flash-sintering. J. Eur. Ceram. Soc. 32, 2293Đ2301 (2012).
[3] Bordia, R. K. and Raj, R. Sintering of TiO2-Al203 Composites: A model experimental Investigation. J . Am. Ceram. Soc. 71, 302Đ310 (1988).
[4] Jha, S. K. and Raj, R. Electric Fields Obviate Constrained Sintering. J. Am. Ceram. Soc. 7, (2014).

E-mail address: Fnu.Shikhar@colorado.edu

DEPARTMENT OF ENGINEERING, UNIVERSITY OF COLORADO, BOULDER, COLORADO

UNFOLDING THE MATHEMATICS ORIGAMI

NOAH WILLIAMS

ABSTRACT. In this paper, we explore how origami is an interface between mathematics and art, and we introduce the Huzita–Justin Axioms for constructing the Origami Numbers. Throughout, we present some results due to Alperin, Hatori, Huzita, Justin, and Lang, and we conclude with a proof due to Hatori that with the Huzita–Justin Axioms, we can solve the general cubic equation over the rational numbers.

1. INTRODUCTION

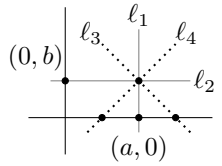
Origami has been fascinating people for hundreds of years, and the the term origami comes from the Japanese words *ori* for folding and *kami* for paper. In its purest form, the art consists of transforming a flat sheet of paper into a sculpture by means of folding alone; no cuts or glue are allowed. Modern origami has evolved in part due to the incorporation of mathematics. In his TED talk, "The Math and Magic of Origami," Robert Lang, physicist and origami master, demonstrates how understanding the mathematical invariants in crease patterns, the blueprints for origami figures, aids modern folders as they invent new figures, sometimes with the help of computer software ([7]). He also shares how origami influences the future of science by providing ways to make objects like space telescope lenses, heart stents, and airbags compact on the way to their destinations. Yet, even as we welcome it into our future, origami motivates a re-investigation of the classical straightedge and compass constructions from the ancient Greeks. The goal of this paper is to demonstrate that origami is an interface between mathematics and art, and in particular, to present some results due to Alperin, Hatori, Huzita, Justin, and Lang. The reader is invited to participate in the folding.

When creating an origami sculpture, one must locate important points on a sheet of paper and then make folds between the points. For simpler designs, it is easy to locate the desired points by folding the paper in half or in fourths. In more complicated constructions, however, especially those that incorporate asymmetry (see Lang's Fiddler Crab, opus 446 in [6]), it is necessary to divide the paper into more difficult ratios that require constructions like trisecting an angle. A mathematical question that one might pose is the following: if we view the complex plane as an infinite sheet of paper and are given the two points $(0, 0)$ and $(1, 0)$, what lines and points can we construct as creases of single folds and their intersections?

At once, we realize that we can fold the x- and y-axes, and that if we have constructed a point $(a, 0)$, then, we can construct points such as $(a, a), (0, a)$, and $(-a, 0)$ by making creases perpendicular to each other through appropriate points. Consequently, it is reasonable to say that the number a is *constructible* if $(a, 0)$ is the intersection of two

Date: February 18, 2015.
Key words and phrases. Math, Origami.

creases formed by single folds. As an example, we show that if a and b are constructible numbers, then so are $a+b$ and $a-b$. Suppose $a, b \neq 0$ (otherwise, there is nothing to prove). First, fold the line ℓ_1 perpendicular to the x-axis through the point $(a, 0)$ and the line ℓ_2 perpendicular to the y-axis through $(0, b)$. The intersection of lines ℓ_1 and ℓ_2 is the point (a, b). Next, make a fold through (a, b) that takes $(a, 0)$ onto ℓ_2. There are two possibilities; call the creases of these lines ℓ_3 and ℓ_4. The intersections of ℓ_3 and ℓ_4 with the x-axis are the points $(a \pm b, 0)$. Consider the following diagram:



.

By repeatedly folding the paper in half, we can construct fractions of the form $\frac{m}{2^n}$, where m, n are non-negative integers, and from this, we can construct arbitrary rational numbers. Indeed, to construct a/b, where b is nonzero, write

$$\frac{a}{b} = \frac{\frac{a}{2^n}}{\frac{b}{2^n}},$$

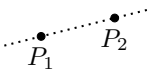
where 2^n is larger than a and b. Now, construct the points $(0, a/2^n)$ and $(b/2^n, 0)$ and fold the line ℓ connecting them. This line has slope a/b. Fold a line ℓ' through the origin that is parallel to ℓ by making two perpendicular folds. The line ℓ' intersects the constructible line $x = 1$ at the point $(1, a/b)$. It is now a simple matter to construct $(a/b, 0)$.

The discussion above motivates a mathematical collection of numbers known as the *origami numbers*, which are constructed on an infinite sheet of paper as intersections of single-creases formed with respect to a set of seven axioms.

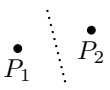
2. The Axioms of Origami

Jacques Justin ([4]) and Humiaki Huzita ([3]) established the following seven axioms that describe how to make a single crease by aligning points and lines on an infinite sheet of paper:

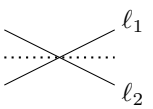
($\mathcal{O}1$) Given points P_1 and P_2, we can make a fold that passes through both.



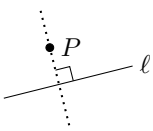
($\mathcal{O}2$) Given points P_1 and P_2, we can fold P_1 onto P_2 to create the perpendicular bisector of segment $\overline{P_1P_2}$.



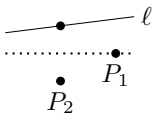
($\mathcal{O}3$) Given two lines ℓ_1 and ℓ_2, we can fold ℓ_1 onto ℓ_2. This is equivalent to finding the angle bisector of the vertical angles formed by ℓ_1 and ℓ_2 as long as these lines are not parallel to each other.



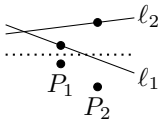
($\mathcal{O}4$) Given a point P and a line ℓ, we can make a fold that is perpendicular to ℓ and that passes through P.



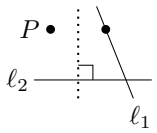
($\mathcal{O}5$) Given points P_1 and P_2 and a line ℓ, we can make a fold that passes through P_1 and places P_2 onto ℓ. This is equivalent to finding the intersection of a line with a circle; there can be 0, 1, or 2 ways to do this.



($\mathcal{O}6$) Given points P_1 and P_2 and lines ℓ_1 and ℓ_2, we can make a fold that places P_1 onto ℓ_1 and P_2 onto ℓ_2. This is sometimes called the neusis axiom.



($\mathcal{O}7$) (Justin–1989 [4], Hatori–2001 [5]) Given a point P and lines ℓ_1 and ℓ_2, we can make a fold that is perpendicular to ℓ_2 and that places P onto ℓ_1.



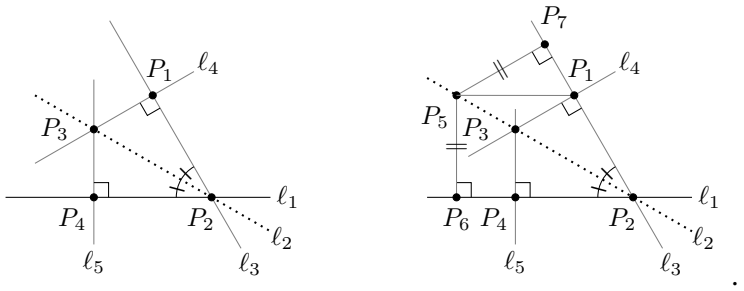
Robert Lang proved that these axioms are a complete list of the possible folds that one can make by aligning a finite number of points and line segments in a single-crease ([2]). In, [1], Roger Alperin builds the origami numbers from the ground up, discussing some of the intermediate fields that can be constructed along the way. (In his paper, the axioms are numbered differently than above.) He shows that axioms ($\mathcal{O}1$) – ($\mathcal{O}4$) give rise to the Pythagorean numbers that constitute the smallest sub-field of $\mathbb{C}$ that contains $\mathbb{Q}$ and that is closed under the operation $a \mapsto \sqrt{1+a^2}$. He demonstrates that if we add ($\mathcal{O}5$), we obtain the Euclidean numbers, which are the numbers constructible via compass and straightedge. Finally, he proves that including ($\mathcal{O}6$) results in the *origami constructible numbers*, the smallest sub-field of the complex numbers that is closed under square roots, cube roots, and complex conjugation ([1]). Interestingly enough, in [5], Koshiro Hatori shows that the origami constructible numbers can be constructed using only ($\mathcal{O}6$). This means that adding ($\mathcal{O}7$), though it is not equivalent to the other folds, does not allow us to construct any more points.

3. SOLVING THE CUBIC

With inspiration from the method presented by Koshiro Hatori in [5], we prove that we can construct a real solution to an arbitrary cubic with rational coefficients. First, we establish the following lemma.

Lemma 3.1. *Given a line ℓ_1 and a point P_1 not on ℓ_1, any fold, ℓ_2, that takes P_1 onto ℓ_1 is tangent to the parabola with focus P_1 and directrix ℓ_1. Furthermore, the point of tangency is constructible.*

Proof. Consider the following diagrams:

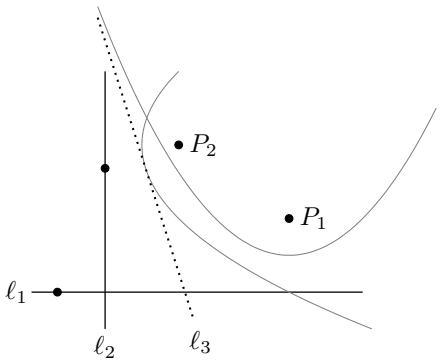


If ℓ_2 is parallel to ℓ_1, we are done, so assume ℓ_2 intersects ℓ_1 at the point P_2. Fold ℓ_3 through points P_1 and P_2 by axiom ($\mathcal{O}1$). Using axiom ($\mathcal{O}4$), make a fold, ℓ_4, perpendicular to ℓ_3 and passing through P_1. This line necessarily intersects ℓ_2 because $\ell_2 \nparallel \ell_1$; call the point of intersection P_3. Now, by axiom ($\mathcal{O}4$), make a fold, ℓ_5, perpendicular to ℓ_1 and passing through P_3. Let P_4 be the intersection of lines ℓ_1 and ℓ_5 (which must exist because $\ell_2 \nparallel \ell_1$). Via the angle-angle-side congruence postulate for triangles $\Delta P_2P_3P_4$ and $\Delta P_2P_3P_1$, it follows that segments $\overline{P_3P_4}$ and $\overline{P_3P_1}$ have the same length, so P_3, which we have constructed, is on the parabola with focus P_1 and directrix ℓ_1 (note that angles $\angle P_3P_2P_4$ and $\angle P_3P_2P_1$ are congruent by axiom ($\mathcal{O}3$)).

No other point on line ℓ_2 can possibly be on the parabola. Indeed, suppose P_5 on ℓ_2 is on the parabola. Then, segment $\overline{P_1P_5}$ has length greater than or equal to the length of segment $\overline{P_5P_6}$ since $\Delta P_1P_5P_7$ is a right triangle with hypotenuse $\overline{P_1P_5}$. (Similar reasoning to that above shows that $\overline{P_5P_6}$ and $\overline{P_5P_7}$ have the same length.) □

Theorem 3.2 (Justin [4])**.** *We can construct a real root of* $p(x) = x^3 + ax^2 + bx + c \in \mathbb{Q}[x]$ *using the axioms of origami.*

Proof. (Hatori [5]) If $c = 0$, then 0 is a root of $p(x)$, and there is nothing to prove, so we assume $c \neq 0$. Recall from our discussion above that we can construct the rational numbers. It follows that we can construct the points $P_1 := (a, 1)$, $P_2 := (c, b)$. We can also fold the lines $\ell_1 : y = -1$ and $\ell_2 : x = -c$. By axiom ($\mathcal{O}6$), we can make a fold that places P_1 onto ℓ_1 and P_2 onto ℓ_2. Let $\ell_3 : y = tx + u$ be the line of the crease. Then, we claim that t is a real root of $p(x)$. Consider the following picture:



By Lemma 3.1, we know that ℓ_3 is tangent to the parabolas $\mathcal{P}_1$ and $\mathcal{P}_2$, where $\mathcal{P}_i$ has focus P_i and directrix ℓ_i for $i = 1, 2$. Furthermore, we can construct, (x_1, y_1) and (x_2, y_2), the points where ℓ_3 intersects $\mathcal{P}_1$ and $\mathcal{P}_2$, respectively. One can verify that the

equations of the parabolas are

$$y = \frac{1}{4}(x-a)^2 \quad \text{and} \quad x = \frac{1}{4c}(y-b)^2. \tag{3.1}$$

The derivatives of these equations at (x_1, y_1) and (x_2, y_2), respectively, are

$$\left.\frac{dy}{dx}\right|_{(x_1,y_1)} = \frac{x_1 - a}{2} \quad \text{and} \quad \left.\frac{dy}{dx}\right|_{(x_2,y_2)} = \frac{2c}{y_2 - b},$$

and both are equal to t. Hence, we have

$$x_1 = 2t + a \quad \text{and} \quad y_2 = \frac{2c}{t} + b. \tag{3.2}$$

Substituting equations (3.2) into equations (3.1) yields

$$y_1 = t^2 \quad \text{and} \quad x_2 = \frac{c}{t^2}. \tag{3.3}$$

We conclude via equations (3.2) and (3.3), that

$$t = \frac{y_2 - y_1}{x_2 - x_1} = \frac{2c/t + b - t^2}{c/t^2 - (2t + a)},$$

which simplifies to

$$t^3 + at^2 + bt + c = 0.$$

It remains to demonstrate that we can construct t. Indeed, using $(\mathcal{O}4)$ twice, we can construct a line, ℓ_3' through $(0,0)$ that has slope t. The intersection of the constructible lines $x = 1$ and ℓ_3' is $(1, t)$. Using $(\mathcal{O}4)$ again, with $(1, t)$ and the y-axis allows us to construct the line $y = t$, which intersects the y-axis at the point $(0, t)$. It follows that t is constructible. □

We conclude by mentioning two important consequences of Theorem 3.2: Using origami constructions, we can double the cube (by solving $x^3 - 2 = 0$) and trisect the angle (by solving $4x^3 - 3x - \cos\theta = 0$, where θ is the given angle). It is well-known that these constructions are impossible using a straightedge and compass alone.

REFERENCES

[1] Alperin, Roger, *A mathematical theory of origami constructions and numbers*, New York Journal of Math. 6 (2000) 119–133.

[2] Alperin, Roger and Robert Lang, *One, two, and multi-fold origami axioms*, Origami 4, Robert J. Lang, ed., A K Peters Ltd. (2009) 371–394.

[3] Humiaki Huzita, *Axiomatic development of origami geometry*, Proceedings of the First International Meeting of Origami Science and Technology, Humiaki Huzita, ed. (1989) 143–158.

[4] Justin, Jacques, *Résolution par le pliage de l'équation du troisième degré et applications géométriques,* Proceedings of the First International Meeting of Origami Science and Technology, Humiaki Huzita ed. (1989) 251–261.

[5] Hatori, Koshiro, *Origami Construction*, K's Origami, http://origami.ousaan.com. Accessed 25 Jan 2015.

[6] Lang, Robert, *Robert Lang Origami*, http://www.langorigami.com. Accessed 25 Jan 2015.

[7] Lang, Robert, *Robert Lang: The math and magic of origami* [video file] (2008). Retrieved from http://www.ted.com/talks/robert_lang_folds_way_new_origami.

E-mail address: Noah.Williams@colorado.edu

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF COLORADO, BOULDER, COLORADO