SCHOOL OF ENGINEERING RESEARCH SHOWCASE 2020

2020 Research Showcase Posters
Table of Contents

BIOENGINEERING

23. Transfer Lithography Optimization for the Development of Microfluidic Wearable Sensors
Shani Williams – BIOE Undergraduate
Advisor: Dr. Emre Araci

CIVIL, ENVIRONMENTAL, AND SUSTAINABLE ENGINEERING

5. Flooring Systems with Prestressed Steel Stringers for Cost Benefit
Alexandra Rivera – MECH Undergraduate
Rafaela Barros Barreto – MECH Undergraduate
Advisor: Dr. Amin Ghafooripour

COMPUTER SCIENCE AND ENGINEERING

30. Causal Modeling for Cybersecurity
Dr. Suchitra Abel – CSEN Postdoctoral Fellow
Jake Singh – CSEN Master's Student
Andrew Tang – CSEN Master's Student
Ethan Paek – CSEN Undergraduate
Advisor: Dr. Ahmed Amer

29. SBChain: A Parameterized, Hierarchical Blockchain
Daniel Okazaki – CSEN Master's Student
Stephen Pacwa – CSEN Master's Student
Advisors: Dr. Ahmed Amer and Dr. Ahmed Ezzat

24. Fog Development Kit: A Platform for the Development and Management of Fog Systems
Christopher Desiniotis – CSEN Undergraduate
Colton Powell – CSEN Master's Student
Advisor: Dr. Behnam Dezfouli

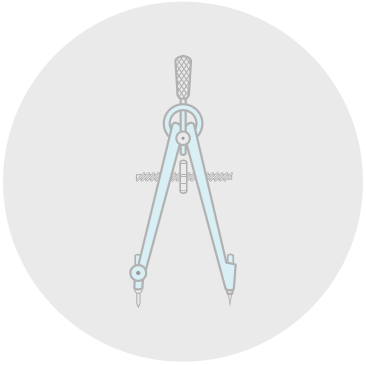
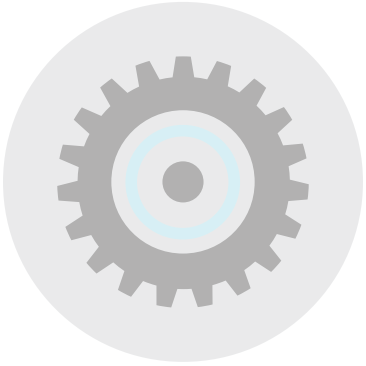
37. Deep Learning for Block-level Compressive Video Sensing
Yifei Pei – CSEN Master's Student
Advisors: Dr. Ying Liu and Dr. Nam Ling

38. A Super-Fast Deep Network for Moving Object Detection
Bingxin Hou – CSEN Ph.D. Student
Advisors: Dr. Ying Liu and Dr. Nam Ling

25. EAPS: Edge-Assisted Predictive Scheduling in WIFI-based IoT Networks
Jaykumar Sheth – CSEN Ph.D. Student
Cyrus Miremadi – ECEN Undergraduate
Advisor: Dr. Behnam Dezfouli

35. Lossy Image Compression Encoding Based on Wavelet Transform and Bit Plane Coding
Cihan Ruan – CSEN Ph.D. Student
Advisor: Dr. Nam Ling



36. Sparse Coding of Intra Luma Prediction Residuals for Screen Content Coding
Michael Schimpf – CSEN Ph.D. Student
Advisor: Dr. Nam Ling

39. Channel-Separable Neural Network for Block-based Image Compression
Pengli Du – CSEN Ph.D. Student
Advisor: Dr. Ying Liu

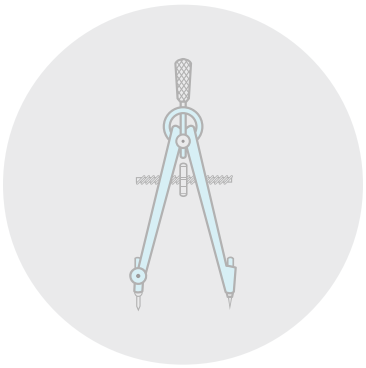
27. A Novel Deep-Learning Approach to Encode Location Information and Its Application On Bike Share Service
Yuan Wang – CSEN Master's Student
Advisor: Dr. Yi Fang

28. Multi-label Classification for Short Text Utilizing the Label Correlations
Zhiyuan Peng – CSEN Master's Student
Advisor: Dr. Yi Fang

26. Modeling the Dynamics of Personal Expertise
Xuyang Wu – CSEN Ph.D. Student
Archana Godavarthy – CSEN Ph.D. Student
Advisor: Dr. Yi Fang

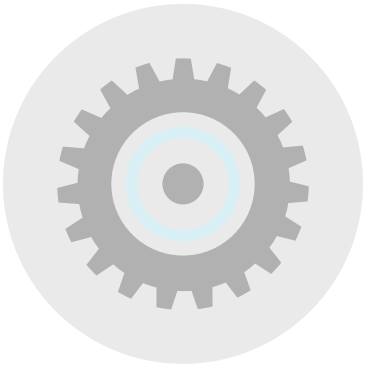
34. Hydration Automation
Peter Ferguson – CSEN Undergraduate
Nicholas Kniveton – CSEN Undergraduate
Jesse Mayer – CSEN Master's Student
Will Tuttle – CSEN Undergraduate
Advisor: Dr. Navid Shaghaghi

ELECTRICAL AND COMPUTER ENGINEERING



20. 3D Nanocarbon Interconnects
Yu Zheng – ECEN Master's Student
Parth Shah – ECEN Master's Student
Dongmeng Li – ECEN Master's Student
Advisor: Dr. Cary Y. Yang

16. Design of Metamaterial Impedance Matching Surfaces at Near Field for EMC Solutions
Ali Khoshniat – ECEN Ph.D. Student
Advisor: Dr. Ramesh Abhari



18. Efficient Hardware/Software Co-design Partitioning of WLANs for Software-defined Radio
Rami Akeela – ECEN Ph.D. Student
Advisor: Dr. Shoba Krishnan

19. Optimal Energy Management of a University Campus Microgrid Integrating Solar Energy and Fuel Cell
Johann Espinosa – ECEN Master's Student
Sowmya Bellam – ECEN Master's Student
Advisor: Dr. Maryam Khanbaghi

15. Reconfigurable Yagi-Uda Antenna Arrays with Passive Beamforming for Microwave Hyperthermia Applications
Nivedita Parthasarathy – ECEN Ph.D. Student
Advisor: Dr. Ramesh Abhari

14. Autonomous Multi-Robot Exploration of Planetary Caves
Kamak Ebadi – ECEN Ph.D. Student
Advisor: Dr. Sally Wood

8. A Deep Learning-enabled Guide for the Visually Impaired
Allen Shelton – ECEN Undergraduate
Advisor: Dr. Tokunbo Ogunfunmi

17. An Input Power-Aware Efficiency Tracking Technique for Energy Harvesters in IoT
Sanad Kawar – ECEN Ph.D. Student
Advisor: Dr. Shoba Krishnan

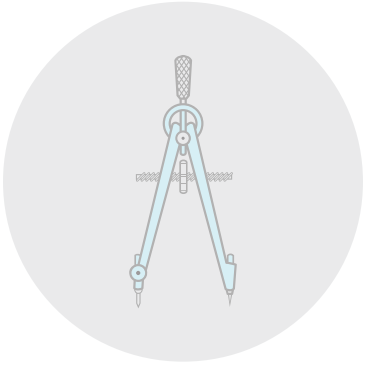
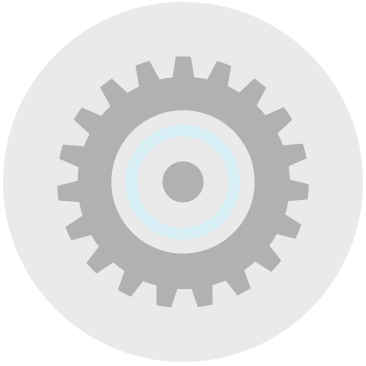
6. Distortion Analysis for a Spherical Transmitter
Amritpal Singh – ECEN Undergraduate
Advisor: Dr. Kurt Schab

9. Face Recognition using Siamese Neural Network
Keshav Pawar – EMGT Master's Student
Advisor: Dr. Tokunbo Ogunfunmi

10. A Fast 2-D Convolution Technique for Deep Neural Networks
Anaam Ansari – ECEN Ph.D. Student
Advisor: Dr. Tokunbo Ogunfunmi

7. Reconfigurable Substrate Integrated Waveguides Using Liquid Metal Microvasculature
Bradley Shirley – ECEN Undergraduate
Advisor: Dr. Kurt Schab



MECHANICAL ENGINEERING

1. Proof of Concept of Swarm Scalar Field Adaptive Navigation
Shae Hart – MECH Ph.D. Student
Advisor: Dr. Christopher Kitts

2. Exploring 3-Dimensional Environmental Phenomena with Drone Clusters
Robert Lee – MECH Ph.D.
Advisor: Dr. Christopher Kitts

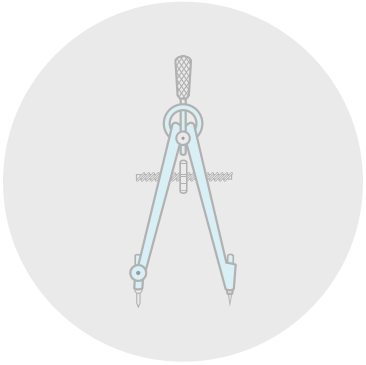
4. Aerodynamics of a Forward-Facing Fin
Tioga Benner – MECH Undergraduate
Advisor: Dr. Godfrey Mungal

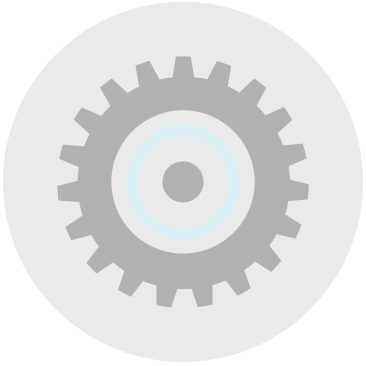
11. Shape Memory Alloy Research for Actuation in Extreme Environments
Rachel Stolzman – MECH Master's Student
Ann McGuire – MECH Master's Student
Advisor: Dr. Christopher Kitts

13. Multi-layer Sensor Data Processing for Occupancy Detection in Residential Buildings
Chenli Wang – MECH Ph.D. Student
Advisor: Dr. Hohyun Lee

12. Development of a Diving AutonomousSurface Vessel and its Applications in Discrete Sampling
Max Woolsey – MECH Ph.D. Student
Advisor: Dr. Christopher Kitts

3. Modeling Bacterial Swimming in Biological Fluids
Noah Lordi – Engineering Physics Undergraduate
Ebru Demir – MECH Postdoctoral Fellow
Advisor: Dr. On Shun Pak





Engineering Stealthy Exosomes for Nanomedicine

Brendan Lawler | Jiayi Zhang | Biao Lu | Santa Clara University | Department of Bioengineering

Abstract

Exosomes are small extracellular vesicles naturally produced by cells for cell to cell communication. Due to their small size, capacity to travel throughout the body, and ability to carry biological cargo such as RNA, DNA, and proteins, exosomes are being studied for potential use in therapeutics.

We have recently developed several mechanisms to modify the exosome membrane, and applied these engineered exosomes towards cell delivery [1,2]. Here, we describe a novel engineering of CD24 exosomes for nanomedicine. CD24 is a glycoprotein that can be found on the surface of some cells such as B lymphocytes. Current research shows that cancer cells also express CD24 to bind to the inhibitory receptor, Siglec-10, on macrophages to evade immune attack [4]. Exosomes, when injected into the body, can be phagocytosed by macrophages. Here we express the novel “do not eat me” signal, CD24, on exosome surfaces to enhance drug delivery efficiency.

We have shown that CD24 can be integrated into a variety of cell types, and have isolated CD 24 exosomes from modified cells. Future uptake experiments will determine if CD 24 is a viable and effective molecule for cell delivery. Due to its stealth abilities and small size, we believe that CD24 has great potential in therapeutics and nanomedicine.

Approach: Surface Engineering

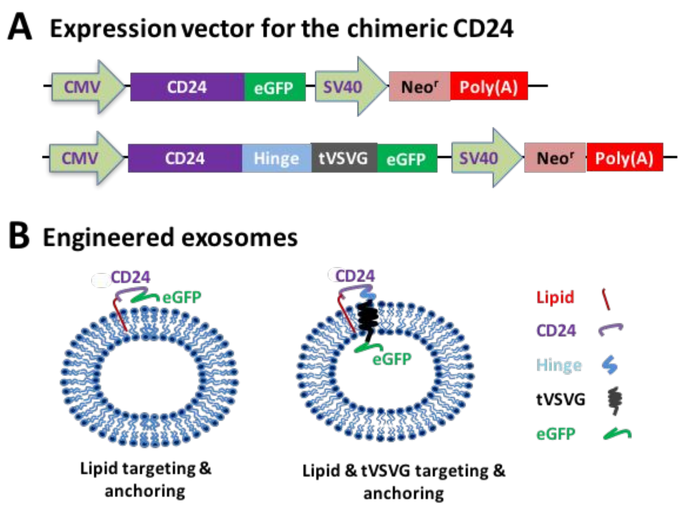


Figure 1: Design and construction of dual-tagged reporter protein for exosome tracking and capture
(A) Vector design for CD24 and tVSVG fusion proteins. Fluorescent (eGFP) genes were attached as shown.
(B) The schematic of loaded exosomes showing CD24 and tVSVG transmembrane protein anchoring with eGFP.

Results

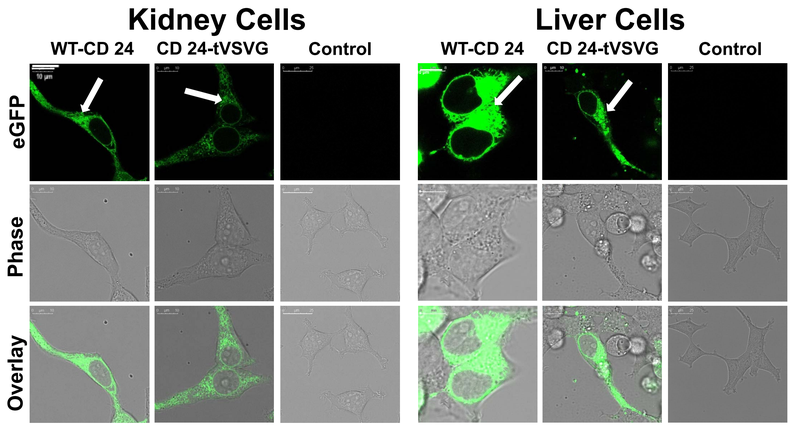


Figure 2: Modified Cell Lines
G2 Cells and 293 cells were transfected with two different constructs to produce CD24 exosomes with eGFP for tracking and imaging.

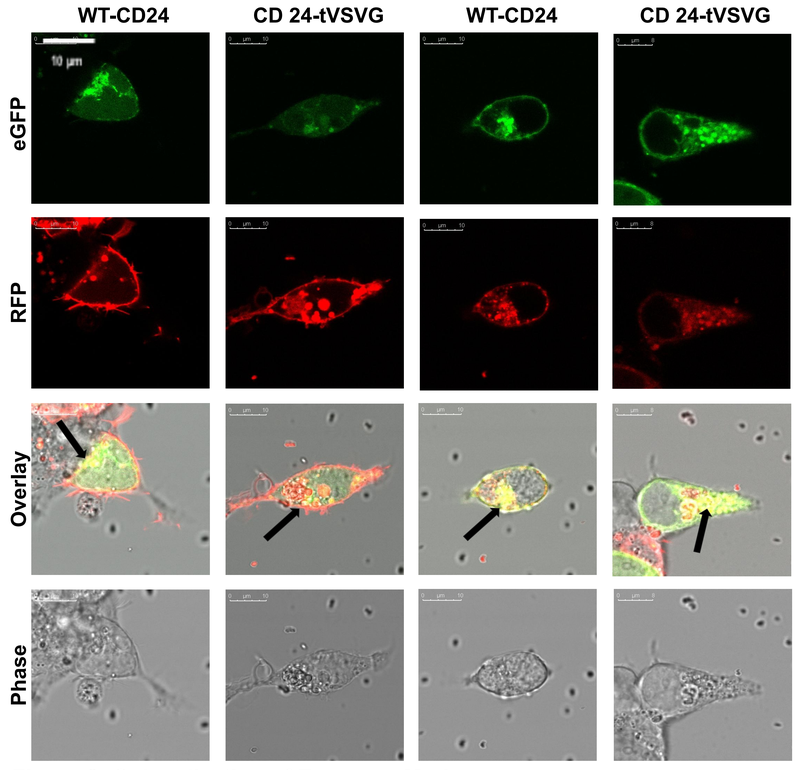


Figure 3: Colocalization of Exosome Markers
293 cells were co-transfected with modified proteins (CD24-eGFP in pc DNA 3.1 or CD24-tVSVG-GFP) and known exosomal marker (WT-CD63). After 48 hours incubation time, exosomes were observed to be colocalized in subcellular locations.

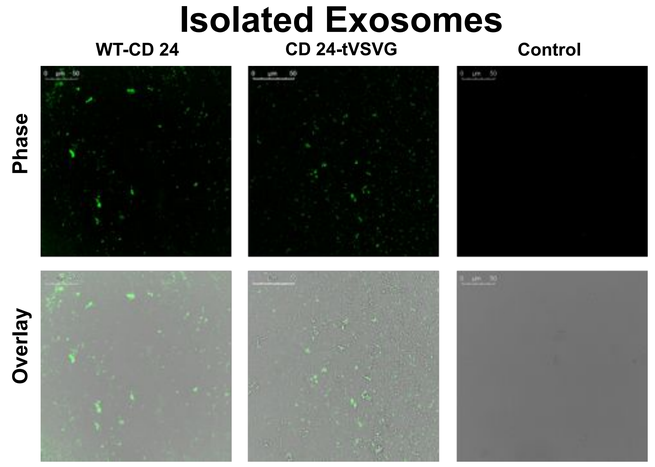


Figure 2: Modified Nanoparticles
Exosomes were isolated from 293T cells and imaged using confocal microscopy.

Summary & Future Work

By using fluorescent imaging, we show that we have successfully integrated CD24 into different cell types and isolated exosomes from modified cells.

Moving forward, we hope to test the stealthy ability of CD 24 exosomes by performing in vitro experiments with macrophages and in vivo experiments using animal models. We also plan to compare CD24 exosomes with exosomes with other stealthy molecules, such as CD47. We hope to show that CD24, with its novelty and small size, will provide exosomes with a longer life span and potential higher drug delivery efficiency than any other current nanomedicine.

Overall, we are hopeful that, with future work, CD 24 “stealthy” exosomes will be a great tool for therapeutics and drug delivery.

Acknowledgement

We would like to thank our lab manager, Annie Brown, for supporting and helping us during the research process.

References

1. Meyer, C., Losacco, J., Stickney, Z., Li, L., Marriott, G., Lu, B. (2017). International Journal of Nanomedicine. 12:3153-3170
2. Stickney, Z., Losacco, J., McDevitt, S., Zhang, Z., & Lu, B. (2016). Development of Exosome Surface Display Technology in Living Human Cells. Biochemical and Biophysical Research Communications, 472(1), 53–59. http://doi.org/10.1016/j.bbrc.2016.02.058
3. Sangiliyandi Gurunathan, Min-Hee Kang, Muniyandi Jeyaraj, Muhammad Qasim, & Jin-Hoi Kim. (2019). Review of the Isolation, Characterization, Biological Function, and Multifarious Therapeutic Approaches of Exosomes. Cells, (4), 307. https://doi-org.libproxy.scu.edu/10.3390/cells8040307
4. Barkal, A. A., Brewer, R. E., Markovic, M., Kowarsky, M., Barkal, S. A., Zaro, B. W., … Weissman, I. L. (2019). CD24 signalling through macrophage Siglec-10 is a target for cancer immunotherapy. Nature, 572(7769), 392-396. Retrieved from https://search-ebscohost-com.libproxy.scu.edu/login.aspx?direct=true&db=edb&AN=138110850&site=eds-live

How do Mammalian Cells Talk to Bacteria Cells

Anna Fraser-Philbin+, Alexander J. Heiler+, Ben Tan, Jonathan Zhang*

Department of Bioengineering, Santa Clara University, Santa Clara, CA 95053

+These authors contributed equally, *PI

Introduction

Many cellular processes are controlled and regulated through cellular communication, both between cells of the same species and between cells from different kingdoms. The mechanisms for mammalian-mammalian, bacteria-bacteria, plant-plant, plant-mammalian, and plant-bacteria communication are all known, leaving bacteria-mammalian communication to be explored.

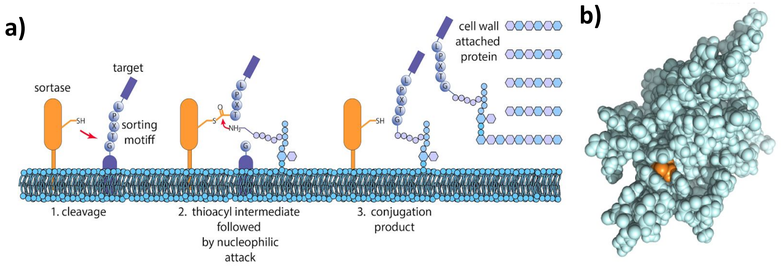


Figure 1. a) Sortase A steps in virulence factor attachment to bacterial cell wall and b) 3D model of Sortase A. a) Sortase A reacts with extracellular targets to form virulence factors, activating bacterial virulence to initiate infection. b) The active site of the enzyme is highlighted in orange.

Prior research in the Zhang bioengineering lab showed the bacterial membrane protein Sortase A (SrtA), which controls the virulence of the bacteria cells, has an active site exuding from the cell surface. We hypothesize the mammalian cells release a messenger molecule to SrtA, controlling bacterial virulence. The purpose of the project is to identify this signal.

Methodology

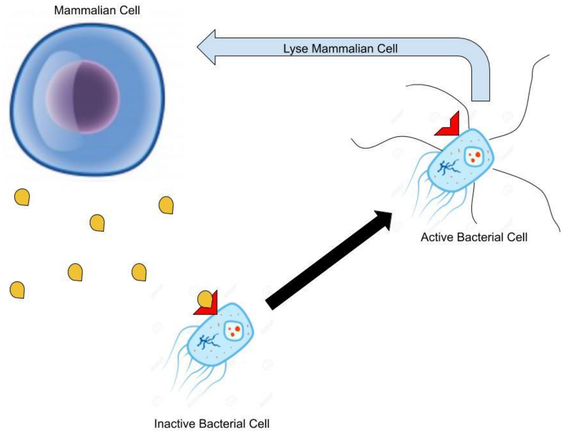


Figure 2. Communication between mammalian and bacterial cells leads to bacterial virulence. Mammalian cells produce and present a ligand that binds to SrtA on the surface of gram positive bacteria. Upon binding, the bacteria cells become activated and infect the mammalian cells.

A non-pathogenic strain of *E. coli* is used to produce the SrtA. The *E. coli* is transfected with the plasmid for Δ59 SrtA 6xHis. This is a variant form of Sortase A with the first 59 amino acids deleted and an added histidine tag on the C-terminus. This engineered form of SrtA is much more stable as the transmembrane region has been removed. Once lysed, the SrtA is captured and purified using the 6xHis tag to perform metal ion affinity chromatography. The purification maintains the conditions for StrA to remain in its active native form.

CHO (Chinese hamster ovary) cells are lysed to extract the entire mammalian proteome. The purified SrtA and the mammalian proteome are then cross-linked using either EDC or benzophenone maleimide as a chemical crosslinker (Fig. 3a and 3c). The resulting protein complexes are again purified using the same affinity chromatography with the 6xHis tag as before.

The cross-linked products are then analyzed by running an SDS-page gel and visualized using silver stain. The gels are compared to find bands shared by all gels. Said bands are then further analyzed using LC-MS-MS (liquid chromatography with tandem mass spectrometry).

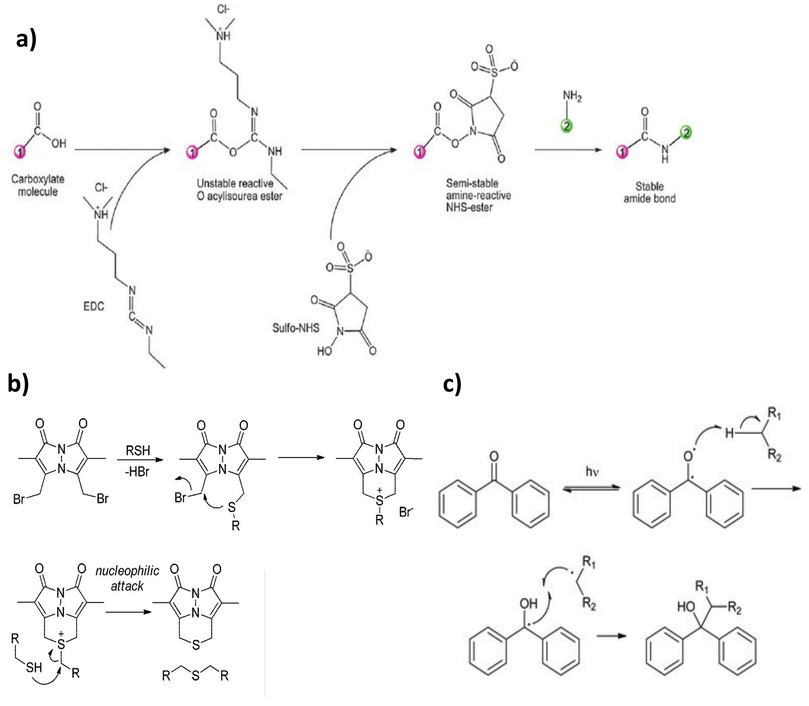


Figure 3. Comparison of chemical crosslinkers.
a) EDC works by nucleophilic attack. It is the weakest of the three shown crosslinkers. b) Dibromobimane works by a nucleophilic attack by a thiol group from the protein on one of the the electrophilic methylbromide groups to generate the thioether. A second nucleophilic attack from the thiol group on the second protein results in the crosslinking of the two proteins. This is stronger than EDC. c) Benzophenone maleimide (BPM) is a photoreactive crosslinker whose mechanism works by inducing a free radical using a UV light source. This is the strongest crosslinker.

Results

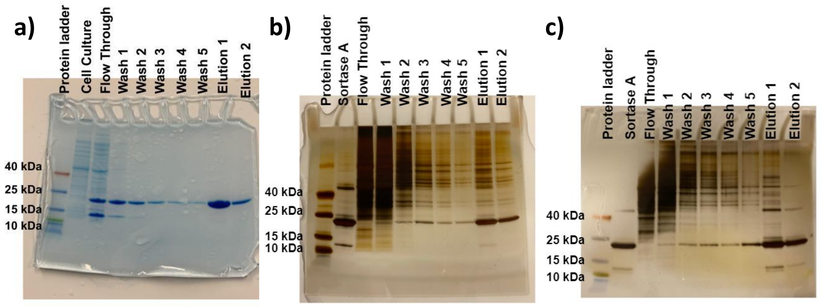


Figure 4. SDS-PAGE of a) Sortase A purification, b) EDC cross-linked products and c) BPM cross-linked products. Each solution was diluted in 12x loading dye and the gels were run at 100V for 1.5 hours. Gel A was stained using a coomassie blue stain and Gels B and C were stained with a silver stain.

Δ59 SrtA with a C-terminal 6xHis tag has a molecular weight of 21.7 kDa (Fig 4a). After crosslinking, the SrtA-signal complex has a molecular weight slightly larger than SrtA, which appears as a band on the gel above the unbound Sortase A band. Due to non-specific binding to the Ni-NTA beads and non-specific crosslinking reactions, byproducts are produced and appear on the gel. To discern the messenger molecule from the byproducts, the gels from each cross-linking trial are overlayed. Due to bio-orthogoniality, the signal molecule should bind to SrtA no matter which cross-linking mechanism is used, while the random byproducts will differ between trials.

Future Steps

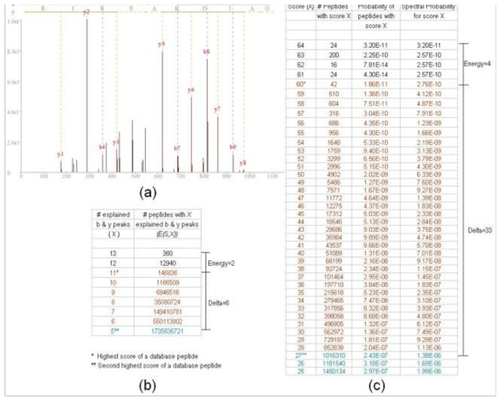


Figure 5. Theoretical data from LC-MS-MS analysis. After trypsin digestion, the cross-linked protein fragments are run through the LC-MS-MS spectrometer. The peaks and fragment molecular weights are compared to a protein database to identify the unknown captured protein bonded to the Sortase A.

Acknowledgements

Funding: SCU School of Engineering Internal Research Grant, Xilinx Grant
SCU Department of Bioengineering: Jerard Madamba
SCU Department of Chemistry and Biochemistry: Owen Gooding
SCU Department of Biology: Daryn Baker

Utilizing a Gaucher's Disease Model for the Evaluation of a Novel Exosome-Based Enzyme Replacement Therapy

Annie Brown | Biao Lu | Santa Clara University | Department of Bioengineering

Abstract

Engineered nano-scale exosomes have great potential as new and targeted delivery vehicles for the treatment of Gaucher's disease, the most common lysosomal storage disease. Recently, we have reported the design, production, and isolation of exosomes loaded with lysosomal β-glucocerebrosidase (GBA). People suffering from Gaucher's disease do not have functional GBA, which results in toxic build-up of undegraded substrates within the cell. To evaluate the efficacy of this exosome-based therapy, a human Gaucher's disease model is required. Here, we have utilized near-haploid human cells (Hap1) modified via CRISPR-Cas9 to model Gaucher's disease *in vitro*. These cells contain a 479 bp insertion in the 6th exon of the GBA gene, resulting in non-functional GBA. PCR, enzyme activity assays, and flow cytometry have been employed to confirm the diseased genotype and phenotype. Characterization of GBA-knock out cells shows a total loss of GBA enzyme activity. Further characterization demonstrates a normal growth rate but an increased number of lysosomes, indicating a diseased phenotype. The utilization of a human GBA-knock out cell line will further enable the evaluation of the efficacy of our engineered exosomes.

Approach: Confirming GBA KO

To disrupt the coding sequence of the GBA gene, a 479 base pair segment of DNA has been inserted in the 6th exon of the gene. This insertion should introduce enough mutations to make the resulting GBA enzyme non-functional.

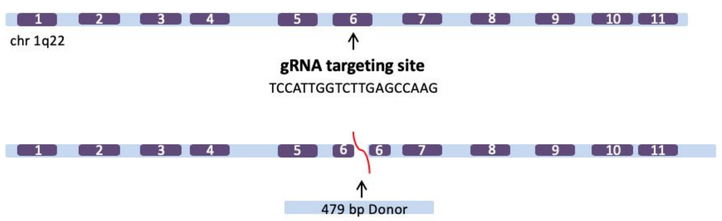


Figure 1. Visual representation of the strategy used to disrupt the GBA coding sequence.

To confirm the knock-out of GBA, the disease model cell line has been analyzed to determine its phenotype and genotype.

Results

First, the morphology of Hap1 control and GBA KO cells was compared. There was no visible difference in the morphology of control Hap1 cells and Hap1 GBA-KO cells. The growth rate of both cell lines was found to be similar, suggesting that the GBA KO cells can be propagated in culture for biologic drug testing.

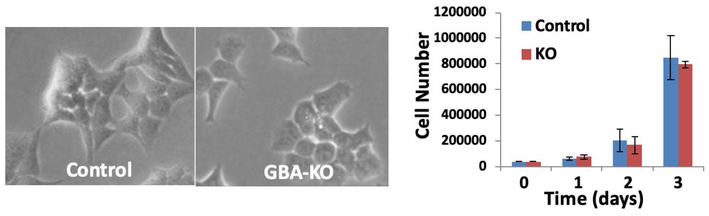


Figure 2. Hap1 GBA KO cells display a similar morphology (left) and growth rate (right) compared to control Hap1 cells.

PCR analysis shows the desired insertion in the GBA gene. An enzyme assay shows extremely decreased levels of GBA activity within the GBA KO cells compared to normal Hap1 cells.

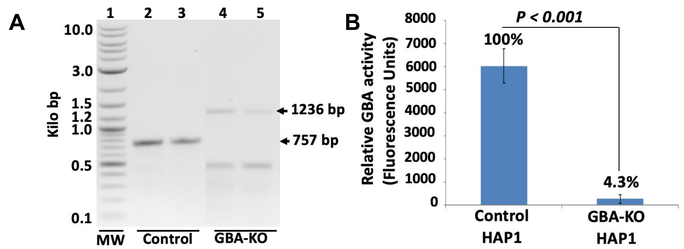


Figure 4. A. The 1236 bp band contains the desired 479 bp insertion. B. Compared to control Hap1, GBA KO cells have less than 10% normal GBA activity.

Hap1 GBA KO cells show a larger aggregation of lysosomes compared to control. This suggests a build-up of undegraded substrates within the cell, similar to what occurs in patients with Gaucher's disease.

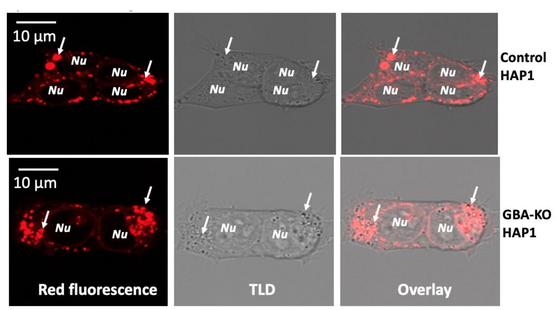


Figure 5. Hap1 KO cells have an increased number of lysosomes (stained red) compared to control.

Initial testing shows that the addition of rhGBA can sufficiently restore the enzyme activity within the GBA KO cells. Three different dosages of rhGBA were tested and the 100 ng and 150 ng doses showed enzyme activity significantly greater than that of the KO cells with no added enzyme.

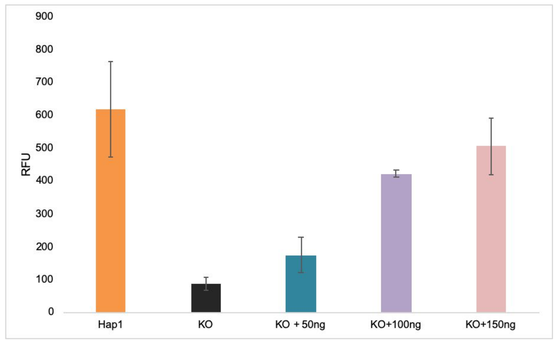


Figure 6. Resultant enzyme activity Hap1 GBA KO cells with the addition of rhGBA compared to WT Hap1. Enzyme activity is proportional to the relative fluorescence detected (RFU). n = 3

Summary & Future Studies

Both the diseased genotype and phenotype of the GBA KO cells has been confirmed through qualitative and quantitative analysis.

Next, we aim to further test the efficacy of GBA enzyme replacement therapies utilizing the diseased model. We have shown the ability to "rescue" GBA KO cells with purified, rhGBA. We also hope to treat the cells with GBA loaded within exosomes and determine whether treating Hap1 GBA KO cells with these exosomes will significantly increase the enzyme activity within the cell.

References

Do, M.A., Levy, D., Brown, A. *et al.* Targeted delivery of lysosomal enzymes to the endocytic compartment in human cells using engineered extracellular vesicles. *Sci Rep* 9, 17274 (2019) doi:10.1038/s41598-019-53844-5

HAP1 knockout cell lines. (n.d.). Retrieved from https://horizondiscovery.com/en/products/gene-editing/cell-line-models/PIFs/Human-HAP1-Knockout-Cell-Lines?nodeid=entrezgene-2629

Santa Clara University
School of Engineering

eVision: Influenza Prediction

Department of Computer Science and Engineering: Andres Calle, Supriya Karishetti, and Prof. Navid Shaghaghi
Department of Bioengineering: George Kouretas

Background

Seasonal influenza, also known as the flu, is a serious respiratory infection caused by influenza viruses. The annual flu epidemics are estimated to result in about 3 to 5 million cases of severe illness, and about 290,000 to 650,000 respiratory deaths worldwide.

Objective

To design a machine learning algorithm capable of taking into account previous epidemic records and Google Search data to make accurate predictions of seasonal influenza patterns. Such a prediction model would be useful to help pharmaceutical companies and health care providers better prepare for the annual flu outbreak.

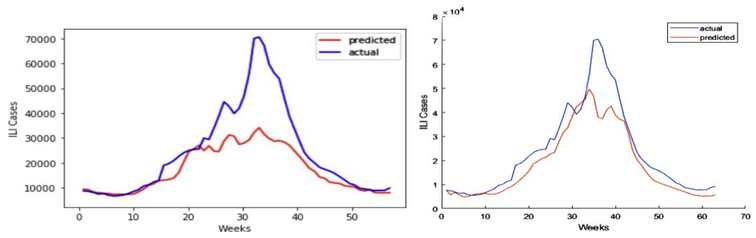
Data Collection

Data was gathered from three major sources, the CDC's FluView portal provided weekly data on influenza-like illnesses reported in the United States since the first week of 1997.

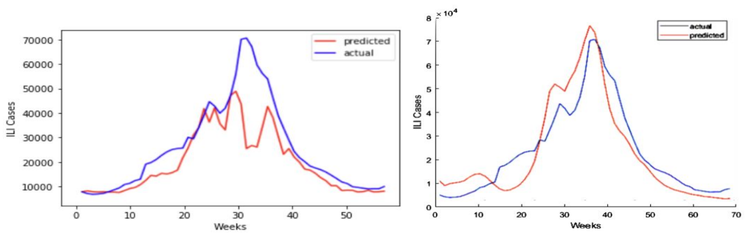
The World Health Organization also began gathering data in 1997, although measuring only the number of influenza viruses analyzed by National Influenza Centers across the world. Data was gathered on four geographically distinct nations: Australia, Chile, the UK, and the US.

Two differing sets of Google Trends data, consisting of four keywords, were gathered to match the CDC and WHO data respectively. Google Trends weekly data is limited to a shifting 5-year window.

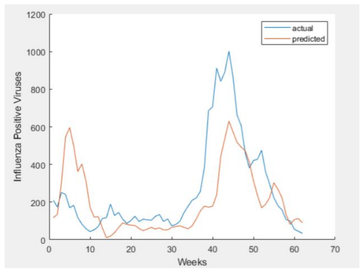
Results



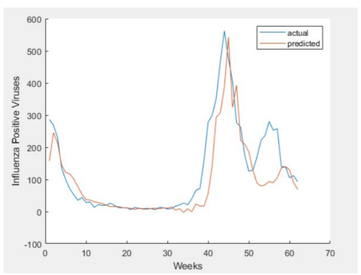
3 Week Influenza Prediction Python vs. MATLAB



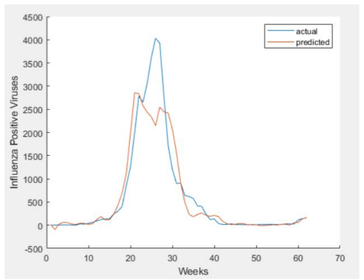
7 Week Influenza Prediction Python vs. MATLAB



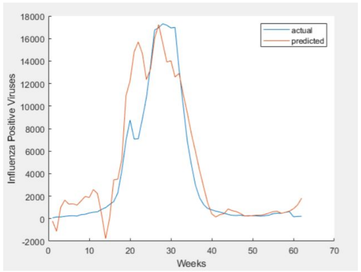
WHO 3 Week MATLAB Australia Prediction



WHO 3 Week MATLAB Chile Prediction



WHO 3 Week MATLAB U.K. Prediction



WHO 3 Week MATLAB U.S.A. Prediction

Program Design

A Long Short Term Memory Neural Network (LSTM) was decided on for the deep learning model due to its utility as a Recurrent Neural Network for multivariate time series forecasting.

Two different models were produced in two languages: Python and MATLAB. The Python model consisted of five total layers while the MATLAB model had four distinct layers.

Conclusion

After designing two competing models in MATLAB and Python, it was observed that the models' accuracy was higher when predicting seven weeks in advance than just three weeks irrespective of the models. Of the two models, the seven week MATLAB model proved to be much more capable at forecasting the season's flu epidemic trend.

Furthermore, the research showed the model's effectiveness when extended to other countries beyond the United States.

Future Work

Current steps for future work will be conducted along two major pathways: expanding granularity and expanding scope. Although the model is built on national data, the CDC and Google Trends break down the data they gather on a state level for the United States.

While the ability to apply the model to nations outside the US has been proven, all predictions occur independently of each other. The scope of the project could be expanded by including multiple nations in a holistic prediction approach.

Peek Into the Molecular Origin of Life – Oligomerization of Peptides Introduce Functions

Carley Fowler[1] , Ben Tan [1] , Blake Williams [1] , Zhiwen J. Zhang [1]

[1]Department of Bioengineering, Santa Clara University

Abstract

There is a knowledge gap in the molecular theory about the origins of life on earth. The RNA World hypothesis states life originated from self-replicating RNA molecules that are the precursors to the proteins and DNA currently studied. However, there is a lack of knowledge on how RNA developed into proteins. This research involving a methodology of Molecular Devolution created a 74 amino acid fluorescent peptide termed iGFP. The study of this peptide suggests that there may have existed a "peptide world" that connects the RNA world to the protein world. In agreement with modern science, this study proves that monomeric peptides do not have structure, and in turn, have no function. However, the formation of a non-covalent peptide complex could produce primitive function as demonstrated in this study. These findings suggest that the simple complexes connected through covalence, such as disulfide bonds and amide bonds, could generate proteins that function "perfectly" in this world.

Introduction

To investigate evolution of peptide oligomerization towards protein structure and its subsequent introduced function, iGFP-fusion protein was used as the model in this study. iGFP fluorescent peptide was fused with three different proteins, glutathione S-transferase (GST), maltose-binding protein (MBP), and small ubiquitin-related modifier (SUMO), separately. GSF protein naturally exists as a dimer, while both MBP and SUMO exist as monomeric proteins. Compared to MBP with the size of 392 amino acids, SUMO protein is small, approximately 100 amino acids in length. Fluorescent intensity of all three fusion proteins were measured and qualitative analyses were carried out.

Experimental Setup

Plasmid preparation: Total of five different DNA plasmids were used in this work, GST-GFPcore, GST-iGFP, MBP-core, MBP-iGFP, and SUMO-iGFP.

Protein purification: Proteins were expressed via bacterial expression system and further purified.

UV absorbance: Qualitative comparisons were carried out both in vitro and in vivo

- GST-iGFP vs. GST-GFPcore
- MBP-core vs. MBP-iGFP
- MBP-iGFP vs. SUMO-iGFP

Results

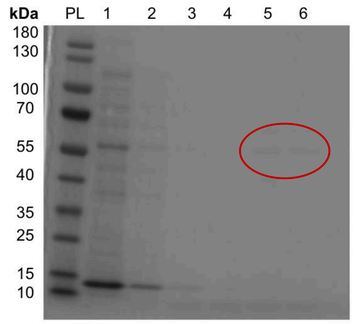


Figure 1. Agarose gel of MBP-iGFP purification. PL is the protein ladder. Lane 1 is the crude extract from the purification process. Lanes 2-4 are the flow-through collection samples. Lanes 5 and 6 are the elution samples from the affinity purification. The red circle is the desired MBP-iGFP protein.

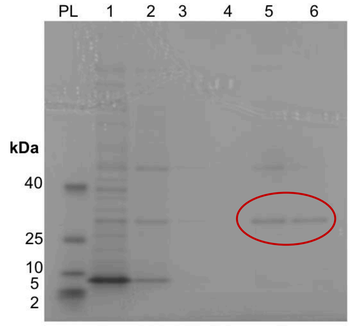


Figure 2. Agarose gel of SUMO-iGFP purification. PL is the protein ladder. Lane 1 is the crude extract from the purification process. Lanes 2-4 are the flow-through collection samples. Lanes 5 and 6 are the elution samples from the affinity purification. The red circle is the desired SUMO-iGFP protein.

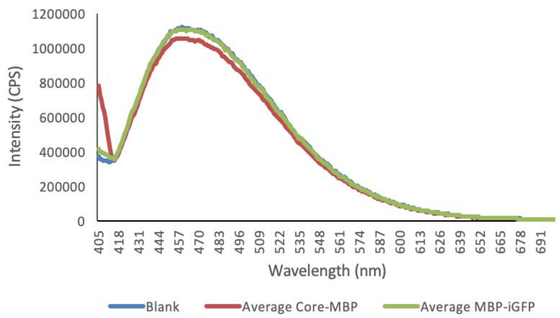


Figure 3. Fluorescence of MBP-iGFP and MBP-Core Fusion Proteins. The fluorescent activity of the fusion proteins showed no significant difference. The monomer protein, MBP-iGFP, had no fluorescence.

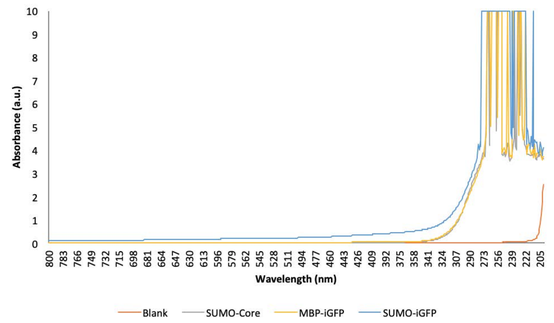


Figure 4. Fluorescence of SUMO-iGFP and SUMO-Core Fusion Proteins. The fluorescent activity of the fusion proteins showed no significant difference. The monomer protein, SUMO-iGFP, had no fluorescence.

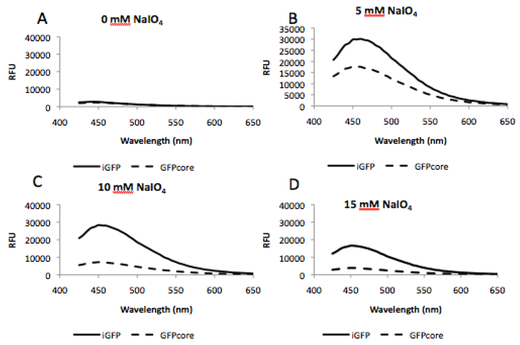


Figure A-D. Fluorescence of GST-iGFP and GST-GFPcore Fusion Proteins. Oxidizing agent NaIO4 activated the iGFP peptide causing fluorescent activity around 450 nm for the dimer protein.

Discussion

The results of this study support the hypotheses of the Peptide World as depicted in Figure 5 below. First, the agarose gels confirmed a successful purification of both the MBP-iGFP and SUMO-iGFP as illustrated in Figures 1 and 2. With the use of a Fluorolog and UV-Vis spectrometer, the monomer proteins showed no fluorescent activity (Figure 3 and Figure 4). However, data collected from the dimer protein, GST-iGFP, showed a florescent response. These results suggest peptide-monomer complexes do not display function; however, peptide-dimer complexes do demonstrate basic functions.

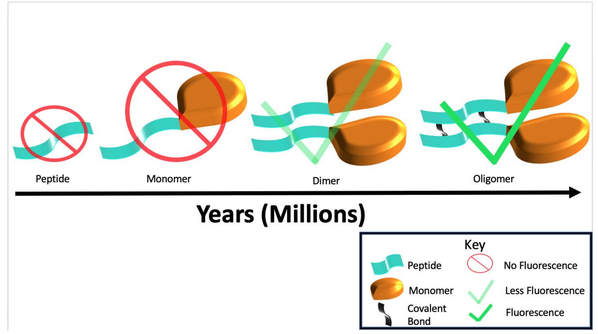


Figure 5. The Peptide World. Peptides do not fluoresce but as the protein increases in complexity, the intensity of the fluorescence also increases. This study suggests a peptide world as the connection between the RNA and Protein World.

Conclusion

1. The lack of fluorescence from the monomer protein suggests that monomeric peptides do, in fact, have no function.
2. The weak fluorescence from the dimer suggests that non-covalent peptide complexes do exhibit primitive functions.
3. This study suggests that a "peptide world" may have facilitated the transition from the RNA World into the Protein World of today.

Acknowledgments

Thank you to Dr. Zhiwen Jonathan Zhang, Blake Williams, and Daryn Baker who made this research possible. Funding was provided by the Kuehler Scholarship and Dr. Zhang's Research Lab.

Contact

Zhiwen Jonathan Zhang
Department of Bioengineering, Santa Clara University
Email: zzhang@scu.edu
Website: https://www.scu.edu/engineering/faculty/zhang-jonathan/
Phone: 408-5546864

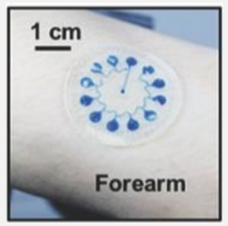
Transfer Lithography Optimization for the Development of Microfluidic Wearable Sensors

Shani Williams, I. Emre Araci
Department of Bioengineering, Santa Clara University

Background

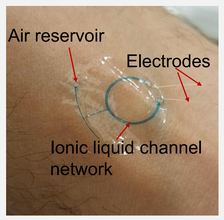
Wearable skin mountable sensors have been created to assist in drug delivery, mprove human-computer interaction, record progression in sports rehabilitation, etc. There are different types of sensors that give these applications.

Sweat sensing[1]
Collects sweat to measure electrolyte balance and biomarker concentration



[1] J. Choi, et al.,., *Adv. Healthcare Mater*, 6/5, 2017

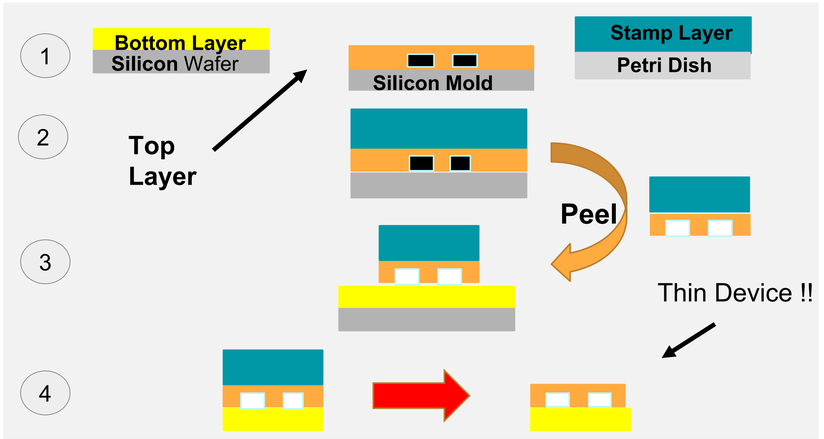
Microfluidic Strain sensing[2]
Uses the change in cross-section or volume of the microfluidic channels to measure strain to track movement



[2] Agaoglu, et. al. *Lab on a chip*, 2018

Problem Statement

STEP 1: Fabrication of thin (top), stamp, and bottom layer
STEP 2: Place the stamp on top of the thin layer and place in oven for 15 mins @ 80°C.
STEP 3: Cut out the combined layer/plasma treat the thin layer to the bottom layer
STEP 4: Remove the stamp from the thin and bottom layer



The goal is to make thin wearable microfluidic devices (~ 50-80 µm) using a modified decal transfer lithography (mDTL).[3] The mDTL utilizes only soft-lithography and reqires optimization of curing times and crosslinker to base ratios to ensure that the temporary bonding between the stamp and thin featured layer is strong enough in **STEP 2**, yet weak enough in **STEP 4**.
[3] W. Childs, et. al. *American Chemical Society*, 124(45), 2002

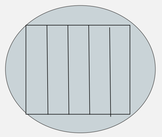
Method

1. Sample Preparation: Soft Lithography

Strips
Mix 1:5 PDMS, spin coat at 500 RPM on silicon wafer
Bulk
Mix 1:20 PDMS to create bulk, pour in 15 cm petri dish

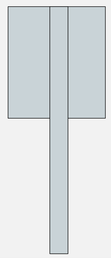
2. Cricut should be used for reliable sample size

PDMS sample to create strips



Cricut

Dimensions of sample



Base
Length: 2.9 cm
Width: 2.5 cm
Thickness: 1 cm
Strip
Length: 7 cm
Width: 1 cm
Thickness: 500 um

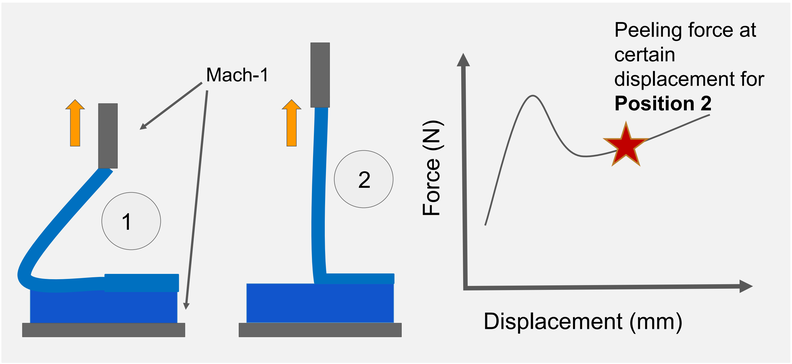
3. Curing and Bonding Times

Curing Times (Strip/Base)
2 hours / 2 hours
2 hours / 1 hour
2 hours / 1.50 hours
2 hours / 45 minutes

Bonding Times
15 minutes
30 minutes
45 minutes
1 hour
24 hours

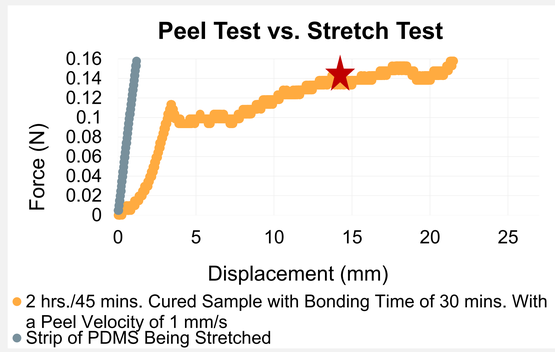
4. MACH-1: 90° Peeling Test[4]

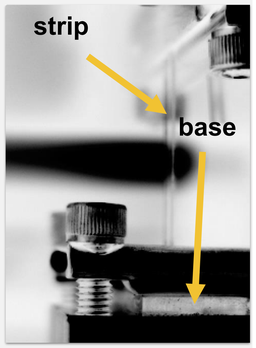
Position 1 shows the strip of the sample before force is applied by MACH-1. As the top fixture moves up in the direction of the arrow, strip peels off.
Position 2 shows strip peeled to 90°
[4] Z. Haifeng, et. al., *Acta Mechanica Solida Sinica,* 21(3), 2008



Results

The slope of the force vs. displacement of a single strip is initially comparable to the slope of the strip during peel test. As soon as the peeling starts there is fast decrease in the force. As the peeling continues, the force gradually increases. The red star indicates the 90° peeling force of 0.145 N. The figure on the right shows peel test of sample at 90°.

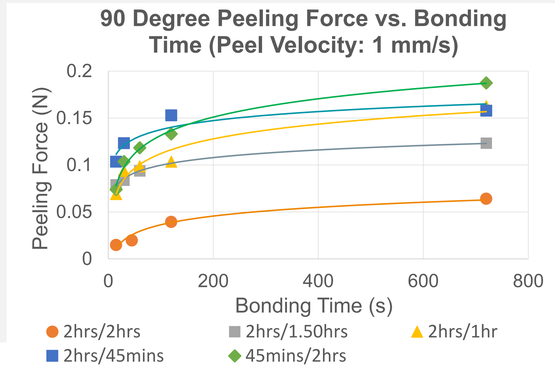


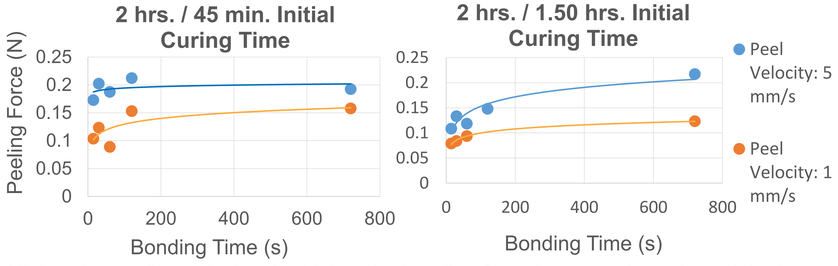


The **longer the bonding time** between the strip and base of the sample **the higher the force** is to peel the strip away from its base.

The **shorter the curing time** of the PDMS in the sample the **higher the force** is to peel the strip away from its base

Note: Solid lines are the logarithmic fit of each data set





Higher the peeling velocity, the higher the bonding force between the strip and the base.

Summary/Conclusion

1. Used MACH-1 to apply force to samples to help identify appropriate curing and bonding times for thin device creation using soft lithography based mDTL.
2. Concluded that even when strip and base is fully cured for 2hrs at 80°C, the off-ratio mixing of the PDMS allows transfer of the thin layers.

Acknowledgements

I would like to thank the Santa Clara University, School of Engineering for awarding me the DeNovo Grant to allow me to participate in the research fellowship this summer.

Flooring Systems with Prestressed Steel Stringers for Cost Benefit

Amin Ghafooripour,[1] Antika Nidhi,[2] Rafaela Barreto,[3] and Alexandra Rivera[4]

Abstract

Steel flooring systems are often overlooked due to their material cost. Prestressing steel and reducing the weight of each element addresses this cost problem. Through a three-part analysis, this research concludes that using prestressed steel for flooring systems is economically viable, when in an optimal configuration. Stage I determined an optimal stringer shape for the flooring system, Stage II focused on finding an appropriate tendon profile while reducing cost, and Stage III continued the cost-benefit analysis to ensure economically beneficial shape and profile. The results show that the prestressed truss with straight tendons has maximum efficiency.

Background

Prestressing creates artificial stresses that would oppose the stresses later applied by a load. The two loads, prestressed and regular, increase the elastic work of the material.

The elastic work is stretched when a prestressing load f_0 is induced in a structure, acting inversely to stress from a load (Figure 1).

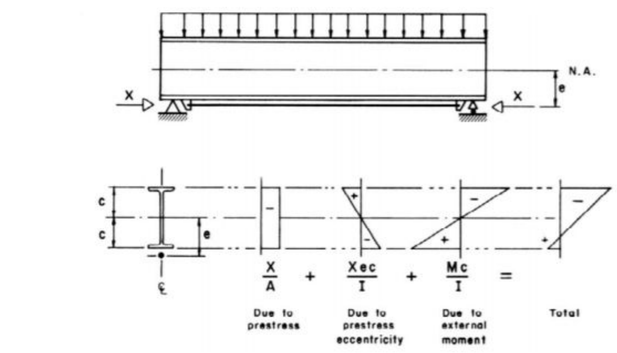


Figure 1: Stress distribution due to prestress and load on a symmetrical steel I-beam

There are three main methods used for prestressing:

Method 1: Placing high-strength steel cables or tendons as far as possible from the neutral axis at the tensile region of the section (Figure 2)

Method 2: Pre-deflecting the beam by deflecting the beam downwards, cambering it upward with a concrete slab, followed by jacking the beam downward again.

Method 3: Applying tension and welding high strength steel plates in the tension zone.

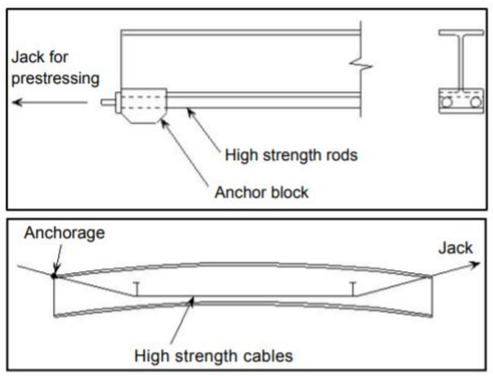


Figure 2: Prestressed beam with tendons

After a review of recent literature on prestressing, there is a noticeable lack of investigation into using prestressed floors for commercial and residential use. Thus, this paper attempts to study not only prestressed I-beam and truss flooring components but also their optimal configuration.

Modeling & Materials

2D models of one steel frame were modeled using STAAD Pro 2007. The floor stringers under investigation were I-section and truss, with spans varying from 5-30m. For each span, the models are prepared for different prestressed loads starting from 0 applied at an interval of 100KN. Figure 3 shows a model of the I-beam stringer and model of the truss stringer.

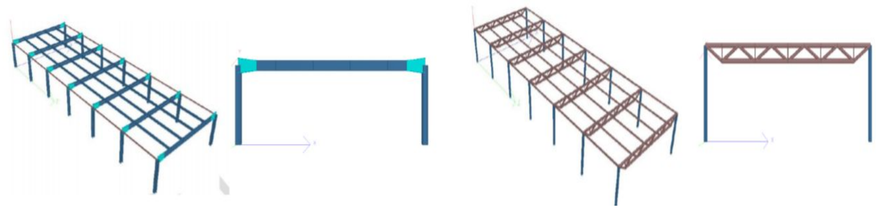


Figure 3: I-beam (left) and truss (right) stringer flooring system

All of the sections used in the frame are standard hot-rolled sections. Universal beams were used for the I-beam stringers and pipes for the truss stringers.

Experimental Process & Results

Stage 1: Determining Optimal Stringer Shape

I-beam Testing: The most suitable condition to investigate the flooring system used horizontal members with straight prestressed loads, applied over the partial span of the frame with uniform eccentricity. All spans sizes (from 5m-30m) were subject to a maximum allowable stress of 275 N/mm^2 and an allowable deflection of the length L divided by 200.

Truss Testing: Trial runs were done on systems using three different types of trusses: Pratt, Warren (vertical chords) and Howe. The Warren truss was chosen as most economical, and following tests were run in the same manner as for the I-beam testing.

The graph below (Figure 4) compares the percent reduction in steel of the two different stringer types. The optimum span for an I-beam frame is 30m; for a truss it is 20m. Each frame can be considered ideal only within a specific range.

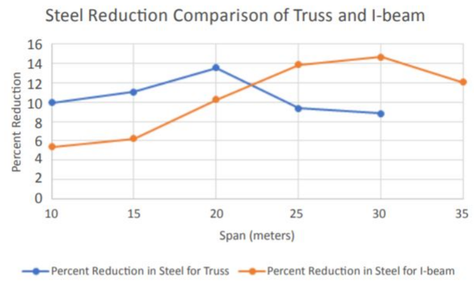


Figure 4: Comparing percent reduction in steel versus span

A second factor to consider when comparing truss and I-beam frames is steel takeoff, used to determine the amount of material needed for purchase and construction. It can be concluded that truss frames are more economical: in all cases, the takeoff is lower (Table 1).

Span	Steel Take off in Kg				Lesser Steel Take off	% Lesser
	I-Beam		Truss			
	Non-prestressed	Prestressed	Non-Prestrssed	Prestressed		
10	1043.5	987.6	889.0	811.1	Truss	17.87
15	2104.8	1973.4	1689.9	1503.0	Truss	23.84
20	3856.3	3460.1	299.5	2589.0	Truss	25.18
25	6100.5	5252.3	4521.4	4041.0	Truss	23.06
30	9392.4	8014.8	6610.1	6026.2	Truss	24.81
35	13885.5	12206.7	-	-	-	-

Table 1: Comparison of steel takeoffs for truss and I-beam frames

Based on Table 1, long spans e.g. 15-30m in a truss frame are more economical than in I-beam frames. With truss frames, steel consumption is 17-25% less.

Stage 2: Optimizing Tendon Profile

A truss frame with a 20m span exhibits maximum steel reduction. To further optimize this flooring system, it is useful to test out different tendon profiles for prestressing in truss frames: straight, curved and V-shaped. The results based on weight reduction indicate that a straight tendon for the entire span, along with prestressing the bottom chord of the truss, is the most economical option.

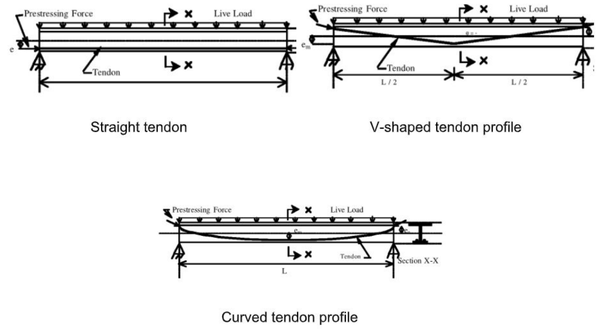


Figure 6: Different type of assumed tendon profiles

Stage 3: Cost Benefit Analysis

With data available from the BCSA website, it was found that the base of installation cost for prestressed steel is £2000 per metric ton. After combining this cost with measurements specific to a 20m span truss frame (best fit for prestressing), it can be seen that a prestressed truss flooring allows for a 7.6% decrease in price. From this brief cost benefit analysis, it can be concluded that the prestressed flooring system has potential for economic and commercial viability.

Conclusion

The results show that a prestressed flooring system could be economically viable for constructing buildings. The following observations can be drawn:

- I-beams stringer: 30m span is optimal, with max steel reduction of 18.78%.
- Truss stringer: 20m span is optimal, with max steel reduction of 13.58%.
- I-sections versus truss: for the same span (ranging 10-30m), in each case truss proved to be more economical, with a lower steel takeoff of 17-25%.
- Straight tendon is the most economical, with steel reduction of 13.58%.
- Prestressing reduces stringer deflection when compared to the non-prestressed frames. Numerical analysis gives nearly the same result as the experimental data, with a maximum deviation of 2.23%.

A flooring system is proposed for a span 15-25m with prestressed truss as main stringer, prestressed over the entire span with straight tendons. This system will have a 10-13% steel reduction. A flooring system with prestressed steel is economically viable, with a 7.6% cost decrease. The prestressed flooring system for a span of 10-35m with straight tendons has potential to be economically and commercially viable.

Acknowledgments

Some part of this research was conducted at Heriot Watt University through a graduate MSc dissertation. The first and second authors would like to thank the School of the Built Environment at Heriot Watt University (Dubai Campus) for their support.

[1]PhD, PE, MASCE, SEI, Department of Civil Engineering, Santa Clara University, USA [2]MS Civil Engineering, Graduate of the Heriot-Watt University, UK (Dubai Campus) [3]Research Assistant, School of Engineering, Santa Clara University, USA [4]Research Assistant, School of Engineering, Santa Clara University, USA

Causal Modeling for Cybersecurity

Jake Singh, Ethan Paek, Andrew Tang, Dr. Suchitra Abel, Faculty Advisor: Dr. Ahmed Amer
Department of Computer Science and Engineering

Background

This study was motivated by events surrounding the “Cloudbleed” data breach.

In 2017, the internet infrastructure company Cloudflare announced that a security flaw had caused the leakage of customer data. These leaks were occasional and involved only small pieces of information. Eventually, a vulnerability researcher at Google was able to attribute the flaw to a buffer overflow to one of Cloudflare’s services.

Objective

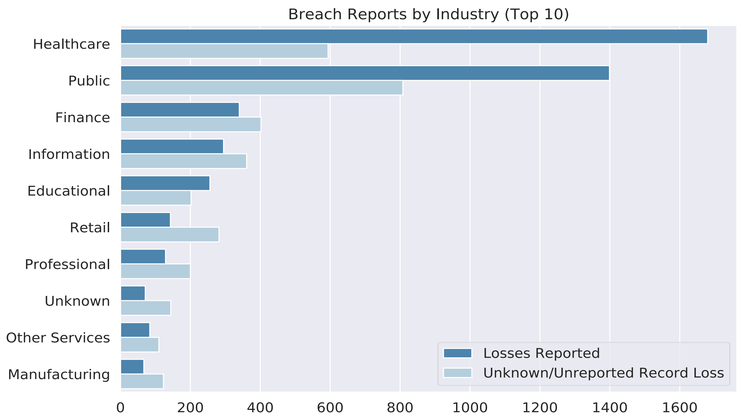
The purpose of this study is to explore a generalization of the "Cloudbleed" case and perform an exploratory analysis in hopes of finding meaningful causal relationships. To do so, we analyze data on real-life cybersecurity incidents using Causal Modeling techniques.

We theorize that actions involving hacking and malware are the most prominent causes of data loss.

Data Acquisition

For identifying potential patterns in our causal model, we initially referred to the annual Data Breach Investigation Reports (**DBIR**) presented by Verizon. The DBIR and the VERIS Community Database (**VCDB**) breaches share a similar framework for categorizing breaches and their attributes.

We chose the VCDB for its greater breadth and more comprehensive features. Data analysis was also performed with DoWhy, Microsoft's open source causal modeling framework.



Research Design

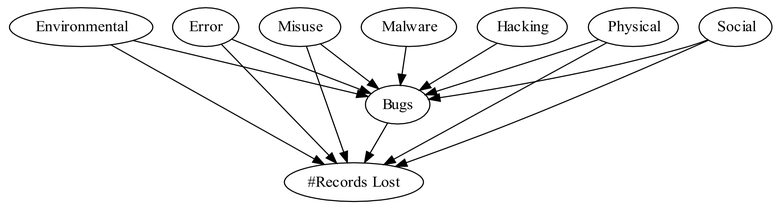
We use the following to determine the **strength** of causation:

- Propensity Score Stratification
- Propensity Score Matching
- Linear Regression Estimator

We rely on **refuters** to show the **robustness** each estimator. This involves rerunning causal analysis methods with the following changes to the dataset:

- *Placebo Treatment*- Replaces the treatment variable with a placebo variable with random values
- *Data Subset Removal*- Runs the program over a randomly chosen subset of the original data
- *Common Cause Addition* - Generates a random confounding variable

Proposed Causal Model



Results

Propensity Score Stratification, and Linear Regression Estimator all result in high causal estimate values (> 1). While these estimators demonstrate a clear causal relationship between action taken and data lost, we theorize that the exaggerated values represent a selection bias within the dataset.

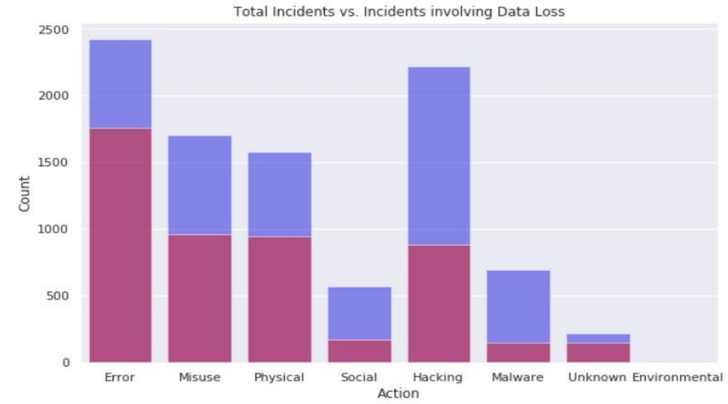
The VCDB provides us with a large selection of data breaches for a wide variety of companies. What the dataset lacks are scenarios where no data breach has occurred, generating an inherent bias within the dataset. This bias makes it so that the data does not fit well into a logistic regression, hence providing an overly large value as the result.

The true strength of the causal model is shown through use of refuters. Placebo, Data Subset, and Common Cause refuters were tested against Propensity Score Matching and Stratification. While refuters act to induce noise in the model, we see stability for each measure of strength. This confirms that a strong causal relation does indeed exist for hacking.

Hierarchical Data in Columnar Form (VCDB)

act-> Hack	act-> hack-> variety-> CSRF	…	act-> Social	act->social->variety- >Spam	…	Records Lost
True	True	…	False	False	…	1024
True	False	…	False	False	…	1
False	False	…	True	True	…	NA
True	True	…	True	False	…	40141

reach ccurrences C



Conclusions

Our program is an exploratory analysis, yielding valuable results. The principal findings of this paper demonstrate a core strength of the causal modeling approach; the ability to use observational data to test hypotheses in a manner that cannot be performed using traditional statistical methods. Because we cannot realistically set up an experiment on data breach incidents, particularly in which all factors are readily provided, DoWhy and Causal Modeling allow us to simulate such experiments and run tests of strength based on data that is missing.

There is still much room for expansion for the use of causal modeling in cybersecurity. We purposefully limited the scope of the factors considered in the Verizon Dataset to actions, leading to a relatively shallow causal graph structure. Creating causal models with additional domain knowledge is a strong candidate for future research.

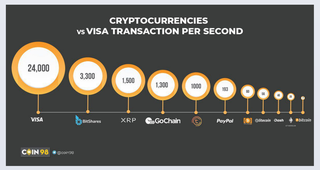
SBChain: A Parameterized, Hierarchical Blockchain

Advisors: Dr. Ahmed Amer Ph.D., Dr. Ahmed Ezzat Ph.D.
Students: Stephen Pacwa and Daniel Okazaki
Department of Computer Science and Engineering

Santa Clara University
School of Engineering

Motivation

- Why use blockchain?
  - Decentralized approach to creating data consensus
  - Allows for a group of non-trusted individuals to create and hold an official record
  - Contains protections against attacks from bad actors
  - Has strong, well-known applications in certain fields; e.g. cryptocurrencies, supply chain management, etc.
- What are blockchains' current problems?
  - Most models currently suffer from scalability issues.
  - These are normally related to speed, throughput, and communication between nodes.



- What is currently being done to solve these issues?
  - Change the structure of the proof system to better accommodate speed and scalability; e.g. proof of work/proof of stake.
  - Do not contain the whole blockchain in one node.
  - Move most transactions off the chain and resolve when necessary; e.g. the Lightning Network.

Objective

- Apply a parameterized, hierarchical blockchain framework to help combat the issues of scale.
  - Build a unified model that can be easily configured to work across a variety of use cases.
  - Build a system that can be configured to work well across a variety of use cases.
  - Parameters can be focused on points like throughput, node trust, percent of nodes required for verification.
  - Communication between nodes is handled in a hierarchical manner.

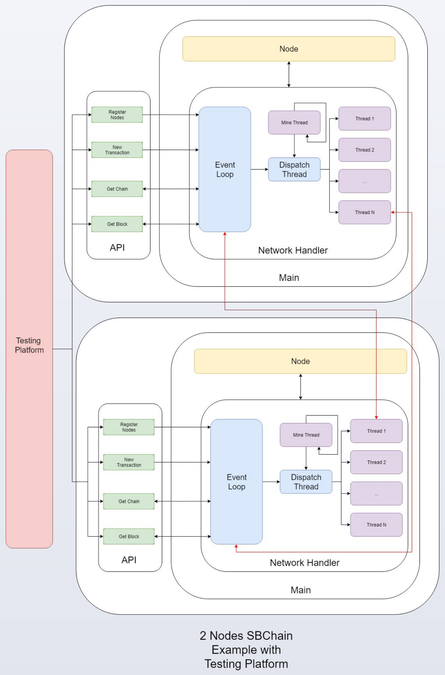
VS

Baseline Design

To determine the efficacy of our new model, we must put it in comparison with an existing representation. Because many traditional models such as cryptocurrencies have very complicated systems in place, they would not be a good match for a simplified prototype of our new model. To solve this problem, we decided to build a simplified model that represents the baseline of what a blockchain would be able to accomplish.

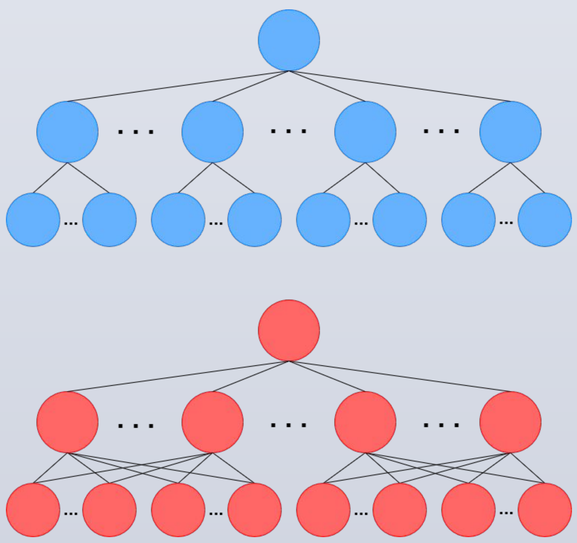
The end goal of this design is to have a system that is capable of completing several tasks through a simple API that can be managed by a testbench. These include:

- Blocks can be mined based on a set difficulty and a proof of work algorithm.
- Transactions can be added to and viewed from the chain.
- Global consensus can be determined by querying local neighbors for information.
- Common abuse cases such as double spending and other attacks are prevented.



2 Nodes SBChain Example with Testing Platform

Hierarchical Design



Our design consists of a simple hierarchical design that is configurable through the use of several parameters. These include aspects such as:

- Height of the hierarchy tree
- Number of node children per parent
- Amount of overlap between children, or the number of 'pools' that a child can be a member of

Currently, there are three models that are being envisioned:

- An element of sharding can be incorporated. Certain levels of the tree can contain subsets of the total ledger with only higher or root nodes containing a complete image of the ledger.
- A similar sharding approach where a partial ledger is kept between all nodes at all levels. The complete image of the ledger can be assembled by viewing nodes in aggregate.
- A model where transactions are triaged to different levels in the tree. Transactions that require more approval or attention can be moved further up in the tree requiring more nodes to view it and reduce the influence that bad actors have.

Applications

There are several applications that would benefit from the use of a hierarchical model. These could include:

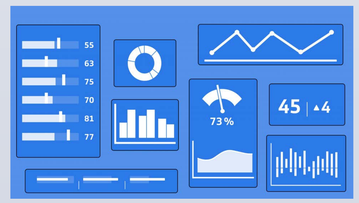
- More scalable and efficient platform for global or finance applications
- Situation where several organizations with competing interests are maintaining a store. The hierarchical mode can prevent one organization from easily overruling the other
- Situation where certain transactions can be approved with different security rates. Lower priority or security transactions can be approved by a smaller group of people
- Education applications for students to learn blockchain

Testing Environment

The testing environment is developed to derive the key performance indicators from both the baseline model and the hierarchical model with several different chosen parameters. It focuses on analyzing points such as:

- Response time
- Transaction latency
- Transaction throughput

This testing environment communicates with the two models through the use of a well-defined API that is kept as similar as possible between the two models. This facilitates the easy transfer of certain test workloads that allow the context of the experiment to remain as similar as possible between experiments.



References

1. https://medium.com/coin98-analytics/compare-the-transaction-speed-of-visa-with-thats-cryptocurrency-ea2fa31d237b
2. https://www.worldcoinindex.com/news/ripple-makes-vis-a-vis-comparison-with-other-cryptocurrencies-for-transaction-costs
3. https://www.pngkey.com/detail/u2w7a9q8q8i1e6u2_speedometer-0shares-car-speed-meter-vector/
4. https://pixabay.com/vectors/search/security/
5. https://www.klipfolio.com/sites/all/themes/klipfolio_theme/img/illustrations/illustration-whatisaKPI.svg?v2

Fog Development Kit:
A Platform for the Development and Management of Fog Systems

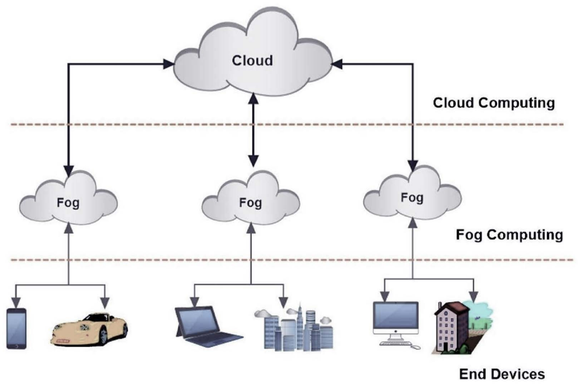
Christopher Desiniotis, Colton Powell and Behnam Dezfouli

Internet of Things Research Lab (*SIOTLAB*), Santa Clara University

Santa Clara University
School of Engineering

Introduction

- By 2030, 500 billion devices will be connected to the Internet [1]
- Cloud computing cannot scale sufficiently to support the processing, storage, and communication demands of billions of IoT devices [2,3]
- Additionally, cloud computing is not suitable for real-time applications
- **Fog computing** aims to address this by bringing more compute, storage, and control capabilities to the network edge



Motivation

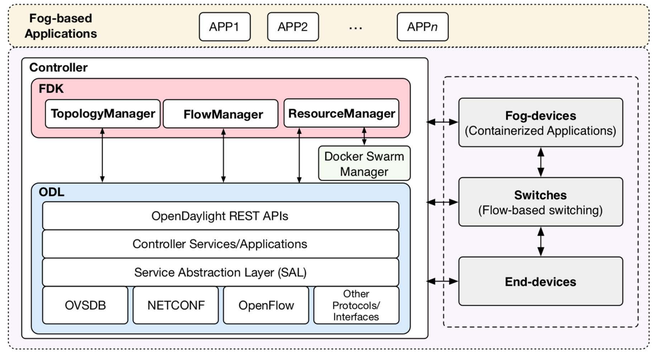
There are significant obstacles for research and development in fog systems:

1. **Complex software components**
   A resource allocator, which reserves both **networking** and **computing** resources, is required in order to allow end-devices to request and reserve resources to meet QoS demands
2. **Costly physical infrastructure**
   A simple fog system requires an infrastructure of end-devices, fog-devices, and networking hardware
3. **Lack of existing tools or platforms**
   Most existing works employ simulation to evaluate efficiency of their resource management approaches [4-6]

Objective

- Design and develop a platform to address the aforementioned challenges
- Abstract the complexity of resource allocation
- Utilize Software Defined Networking (SDN) to reserve network resources
- Utilize virtualization technologies for hosting applications
- Allow for development and deployment of fog-based applications in both physical and emulated environments

System Architecture



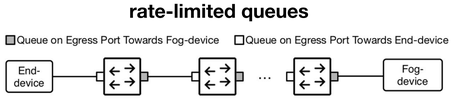
Major Components

- **Controller**
  - Runs OpenDaylight, an SDN controller which enables remote management and configuration of networks
  - Runs the **FDK**, our user-space application that oversees the operation of end-devices, fog-devices, and switches
- **End-devices**
  - Resource-constrained devices that cannot fully satisfy application requirements
- **Fog-devices**
  - To address the processing demands of end-devices
- **Switches**
  - Connect end-devices and fog-devices
  - Support flow-based forwarding

FDK Modules		
Topology Manager	**Flow Manager**	**Resource Manager**
• Query, update, and manage network topology • Provide APIs for managing OpenFlow switches via the OVSDB protocol	• Manage all OpenFlow flows throughout the network • Provide APIs for creating and deleting flows	• Manage and allocate all networking and computing resources • Host server for receiving end-device requests • Provide resource allocation algorithm for reserving resources for an end-device

Resource Allocation

Enforce bandwidth reservation via rate-limited queues

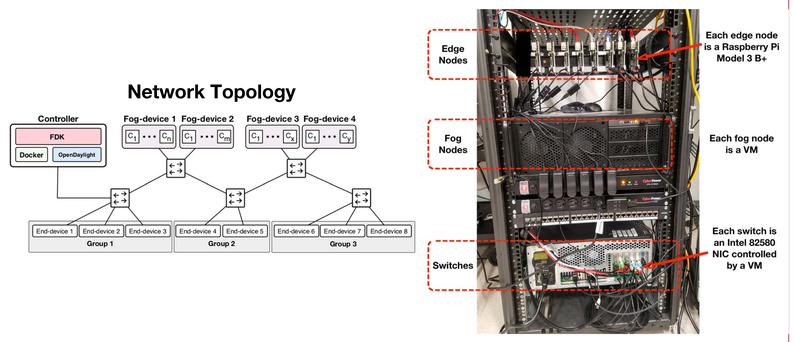


Enforce compute reservation via Docker APIs which utilize cgroups

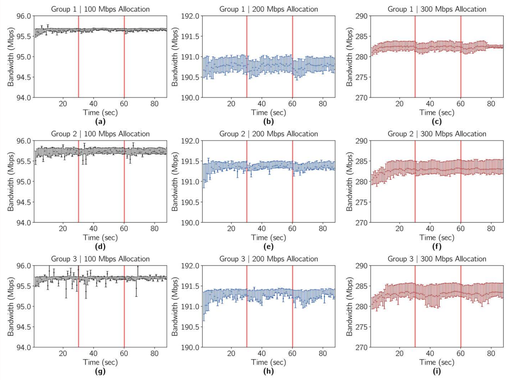
docker

Evaluation

Physical Testbed Used for the Evaluation of the FDK

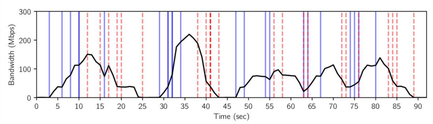


Bandwidth Reservation



- This test demonstrates that the FDK operates well under large volumes of requests and that bandwidth guarantees are reliably fulfilled in a highly congested network
- At times *t=30* and *t=60*, end-device 8 flooded the network with 15 concurrent requests
- Evident by each graph, no application received any significant drop-off in bandwidth

Real-World Image Classification Application



- This test demonstrates a real-world application running on the FDK
- End-devices streamed image data to the fog, via the QUIC transport protocol, for processing by object detection algorithms
- Blue lines denote service requests; red dashed lines denote shutdown requests

References

[1]"Internet of Things, "Cisco.com, 2016. [Online]. Available: https://www.cisco.com/c/dam/en/us/products/collateral/se/internet-of-things/at-a-glance-c45-731471.pdf.
[2] Y. Cao, P. Hou, D. Brown, J. Wang, and S. Chen, "Distributed Analytics and Edge Intelligence: Pervasive Health Monitoring at the Era of Fog Computing," in Proceedings of the Workshop on Mobile Big Data. ACM, 2015, pp. 43–48.
[3] bF. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog Computing and Its Role in the Internet of Things," in Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing, 2012, pp. 13–16.
[4] L. Yin, J. Luo, and H. Luo, "Tasks Scheduling and Resource Allocation in Fog Computing Based on Containers for Smart Manufacture," IEEE Transactions on Industrial Informatics, vol. 14, no. 10, pp. 4712–4721, 2018.
[5] O. Skarlat, S. Schulte, M. Borkowski, and P. Leitner, "Resource Pro61 visioning for IoT Services in the Fog," in IEEE 9th International Conference on Service-Oriented Computing and Applications (SOCA), 2016, pp. 32–39.
[6] J. Son, A. V. Dastjerdi, R. N. Calheiros, X. Ji, Y. Yoon, and R. Buyya, "CloudSimSDN: Modeling and Simulation of Software-Defined Cloud Data Centers," in 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, 2015, pp. 475–484.

Deep Learning for Block-level Compressive Video Sensing

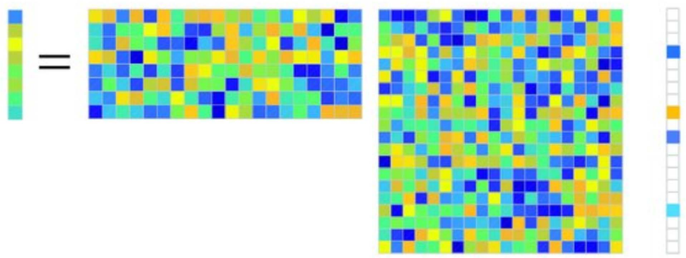
Yifei Pei (M.Sc. COEN student), Dr. Ying Liu, Dr. Nam Ling
Department of Computer Science and Engineering
Santa Clara University

Research Background

- Image compressed sensing: In images, to recover **x** from the measurement **y**, a certain transform is needed, so $\mathbf{x} = \mathbf{\Psi s}$ and **s** is sparse.

$$\underset{\mathbf{s}}{\text{minimize}} \ \|\mathbf{s}\|_1 \quad \text{subject to} \ \mathbf{y} = \mathbf{\Phi\Psi s}.$$

$\mathbf{y} = \mathbf{\Phi} \quad \mathbf{\Psi} \quad \mathbf{s}$



- Traditional methods: Basis pursuit (high computation complexity) and orthogonal matching pursuit (low accuracy).
- Deep learning methods: Not satisfying at very small sampling rates and high computation time at training process.

Objectives

- Introduce the use of discrete cosine transformed images and discrete wavelet transformed images in deep learning for compressed sensing tasks.
- Propose a neural network architecture by combining discrete wavelet transform or discrete cosine transform with deep learning to achieve stronger reconstruction quality of compressed sensed video frames.

Research Design

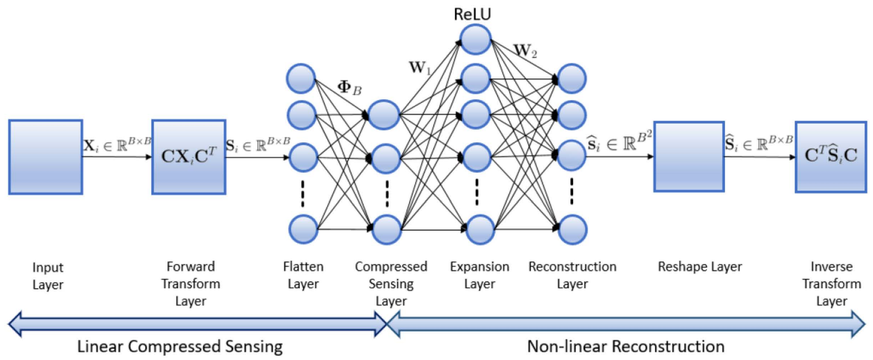


Fig. 1: Fully-connected neural network for compressed sensing.

Experimental Results

TABLE I: The average reconstruction PSNR [dB] versus the sensing rate ($R = M/N$) of the Foreman dataset.

Method	$R = 0.10$	$R = 0.15$	$R = 0.20$	$R = 0.25$	$R = 0.30$	$R = 0.35$	$R = 0.40$
FCN+DCT ($T = 8$)	31.63	32.87	33.87	34.65	35.65	36.34	37.12
FCN+DCT ($T = 10$)	31.55	32.85	34.90	35.52	36.13	36.13	37.33
FCN+DCT ($T = 12$)	31.67	32.80	34.01	34.97	35.70	36.44	37.31
FCN+DWT ($T = 8$)	31.50	32.84	33.79	34.72	35.53	36.10	36.67
FCN+DWT ($T = 10$)	31.49	32.85	33.84	34.61	35.27	36.06	37.11
FCN+DWT ($T = 12$)	31.57	32.84	33.81	34.66	35.49	36.42	36.49
FCN ($T = 8$)	31.22	32.56	33.28	34.18	35.15	35.62	36.65
FCN ($T = 10$)	31.29	32.66	33.35	34.22	35.13	35.75	35.81
FCN ($T = 12$)	31.00	32.39	33.53	34.24	35.02	35.79	36.00
OMP	19.08	20.64	21.85	23.67	24.07	25.11	25.78
BP	20.08	21.60	23.94	25.28	26.55	27.74	28.73
TV	23.91	25.43	27.56	28.83	30.25	31.40	32.24

TABLE II: The average reconstruction PSNR [dB] versus the sensing rate ($R = M/N$) of the Container dataset.

Method	$R = 0.10$	$R = 0.15$	$R = 0.20$	$R = 0.25$	$R = 0.30$	$R = 0.35$	$R = 0.40$
FCN+DCT ($T = 8$)	34.15	35.43	36.56	37.58	38.20	38.87	39.73
FCN+DCT ($T = 10$)	34.20	35.73	36.31	37.23	38.33	39.15	40.27
FCN+DCT ($T = 12$)	34.48	35.64	36.91	37.14	38.44	39.30	39.64
FCN+DWT ($T = 8$)	33.81	35.50	36.33	36.86	37.10	38.06	38.82
FCN+DWT ($T = 10$)	33.92	35.31	36.20	37.05	37.86	37.95	38.83
FCN+DWT ($T = 12$)	34.06	35.29	36.47	37.38	37.51	38.15	38.29
FCN ($T = 8$)	33.70	35.02	35.71	36.24	36.90	37.94	38.78
FCN ($T = 10$)	33.63	34.86	35.36	36.61	36.83	37.94	38.00
FCN ($T = 12$)	34.00	34.74	34.94	36.74	36.75	37.57	38.03
OMP	17.47	18.32	19.22	20.18	20.98	21.77	22.49
BP	18.78	20.19	21.56	22.73	23.73	24.66	25.56
TV	22.33	23.21	24.44	25.47	26.49	27.44	28.28

Fig. 2: Foreman for $M/N = 0.4$. Left to right: original; FCN+DCT ($T = 10$), PSNR = 38.44 dB; FCN+DWT ($T = 10$), PSNR = 38.21 dB; FCN ($T = 10$), PSNR = 36.73 dB; OMP, PSNR = 26.50 dB; BP, PSNR = 29.15 dB; TV, PSNR = 32.90 dB.

Fig. 3: Container for $M/N = 0.2$. Left to right: original; FCN+DCT ($T = 10$), PSNR = 36.97 dB; FCN+DWT ($T = 10$), PSNR = 36.90 dB; FCN ($T = 10$), PSNR = 35.83 dB; OMP, PSNR =19.47 dB; BP, PSNR = 21.48 dB; TV, PSNR = 24.41 dB.

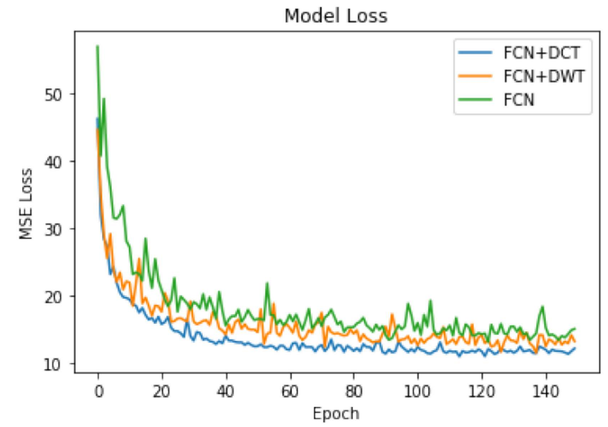


Fig. 4: Validation loss of Container for $M/N = 0.25$ ($T = 10$).

TABLE IV: Total processing time at $R = 0.20$ for 90 testing images (352 ×288).

Method	Time [seconds]
FCN+DCT ($T = 8$)	4.90
FCN+DCT ($T = 10$)	5.12
FCN+DCT ($T = 12$)	5.38
FCN+DWT ($T = 8$)	4.80
FCN+DWT ($T = 10$)	5.01
FCN+DWT ($T = 12$)	5.24
FCN ($T = 8$)	4.13
FCN ($T = 10$)	4.53
FCN ($T = 12$)	4.99
OMP	642.53
BP	543.86
TV	2717.13

Conclusions & Future Plans

- Proposed a deep learning framework that utilizes the sparse property of images to enhance the reconstruction quality of compressed-sensed video frames.
- Demonstrated the sparse transforms can be applied to neural networks to recover compressed-sensed video frames.
- 2D-DCT outperforms 2D-DWT in the FCN.
- Focus on mathematical explanations in the future.

Acknowledgements

This paper has been accepted to *ISCAS 2020*.

A Super-Fast Deep Network for Moving Object Detection

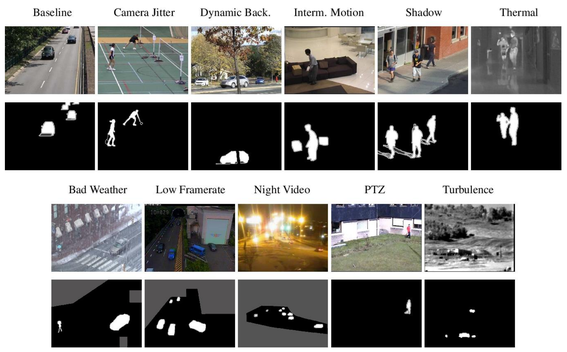
Bingxin Hou (Ph.D. Student), Dr. Ying Liu (Co-Advisor), Dr. Nam Ling (Advisor)

Computer Science and Engineering, Santa Clara University, Santa Clara, CA, 95053

Motivation and Objectives

Motivation:
With the rise of the Internet of Things, the spread of machine vision, and the large amount of video data, it is challenging and crucial to process video data at a fast speed.

Challenge:
Existing deep models in moving object detection are computationally expensive and memory-intensive.



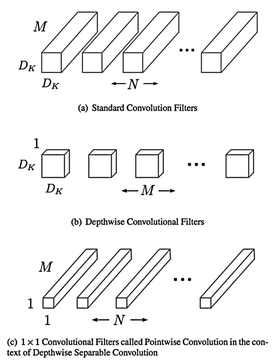
Objectives:
To design a neural network model for moving object detection that does not require explicit background modeling and maintenance, but can significantly accelerate inference speed and still achieve high detection accuracy.

Research Design

Depth-wise Separable Convolutions published in the (Xception paper) or (MobileNet paper). Consist of:

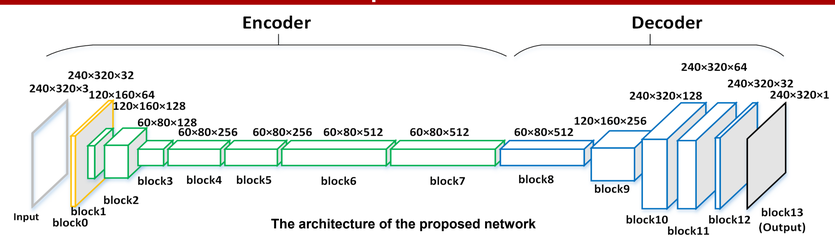
The depth separable convolution replaces traditional convolution by applying an independent 2D filter for each input channel followed by point-wise operation.

The computational cost can be **reduced to $\frac{1}{N}+\frac{1}{D_K^2}$ of a regular convolution**, where N is the number of output channels and D_K is the spatial dimension of the $D_K \times D_K$ kernel.

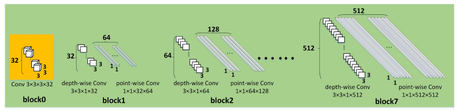


In this paper, we adopt such a separated depth-wise and point-wise convolution structure in a moving object detection network.

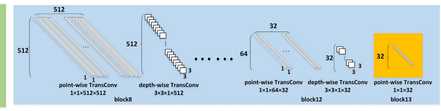
Proposed Model



The architecture of the proposed network



Modified **Depth-wise and point-wise convolution** for Encoder.



Proposed to use **point-wise and depth-wise deconvolution** in decoder phase.

Network Configuration

		Layer Type / Stride	Filter Shape	Output Shape	Parameters #
				240×320×3 (Input Image)	
Encoder	block 0	Conv / s=1	3×3×3×32	240×320×32	896
	block 1	Conv dw / s=2	3×3×1×32 dw	120×160×128	320
		BN		120×160×128	64
		Conv pw/ s=1	1×1×32×64 pw	120×160×64	2048
	block 2	Conv dw / s=1	3×3×1×64 dw	120×160×64	640
		BN		120×160×64	128
		Conv pw/ s=1	1×1×64×128 pw	120×160×128	8192
	block 3	Conv dw / s=2	3×3×1×128 dw	60×80×128	1280
		BN		60×80×128	256
		Conv pw/ s=1	1×1×128×128 pw	60×80×128	16384
	block 4	Conv dw / s=1	3×3×1×128 dw	60×80×128	1280
		BN		60×80×128	256
		Conv pw/ s=1	1×1×128×256 pw	60×80×256	32768
	block 5	Conv dw / s=1	3×3×1×256 dw	60×80×256	2816
		BN		60×80×256	512
		Conv pw/ s=1	1×1×256×256 pw	60×80×256	65536
	block 6	Conv dw / s=1	3×3×1×256 dw	60×80×256	2816
		BN		60×80×256	512
		Conv pw/ s=1	1×1×256×512 pw	60×80×512	131072
	block 7	Conv dw / s=1	3×3×1×512 dw	60×80×512	5632
		BN		60×80×512	1024
		Conv pw/ s=1	1×1×512×512 pw	60×80×512	262144
Decoder	block 8	TransConv pw / s=1	1×1×512×512 pw	60×80×512	262144
		Conv dw / s=1	3×3×1×512 dw	60×80×512	4608
	block 9	TransConv pw/ s=2	1×1×512×256 pw	120×160×256	131328
		Conv dw / s=1	3×3×1×256 dw	120×160×256	2304
	block 10	TransConv pw / s=2	1×1×256×128 pw	240×320×128	32896
		Conv dw / s=1	3×3×1×128 dw	240×320×128	1152
	block 11	TransConv pw / s=1	1×1×128×64 pw	240×320×64	8256
		Conv dw / s=1	3×3×1×64 dw	240×320×64	576
	block 12	TransConv pw/ s=1	1×1×64×32 pw	240×320×32	2080
		Conv dw / s=1	3×3×1×32 dw	240×320×32	288
		TransConv pw/ s=1	1×1×32 pw	240×320×1	33
	block 13	Activation		240×320×1 (Binary Mask)	
Total					983105

For the first time in the literature, we propose the point-wise and depth-wise deconvolution in the decoder network symmetric to the encoder network for computational efficiency.

We propose to use a regular transposed convolution filter and stride 2 for up-sampling as the point-wise transposed convolution, followed by a regular depth-wise convolution to perform as a transposed convolution in depth-wise.

Experiments

Setting

- **Dataset :** CDnet 2014 (15 sequences covering 6 categories were selected)
- **Hardware:** Intel Xeon 8-core 3GHz CPU, Nvidia Titan RTX 24G GPU
- **Software:** PyCharm, Matlab
- **Training frames:** 200 (for each scene)
- **Testing Frames:** 1000+ frames (for each scene)

Hyperparameters

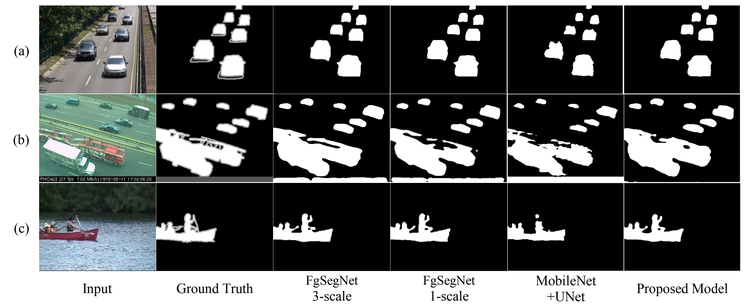
- **Loss function:** Binary Cross Entropy
- **Optimizer:** RMSprop
- **Learning rate:** 10^{-4}
- **Epoch:** 50
- **Batch size:** 1

$\mathrm{F}-measure = 2 \cdot \frac{precision \cdot recall}{precision+recall}$, where $precision = \frac{True\ Positive}{True\ Positive+False\ Positive}$, $recall = \frac{True\ Positive}{True\ Positive+False\ Negative}$..

Results and Performance

Overall comparison with other models

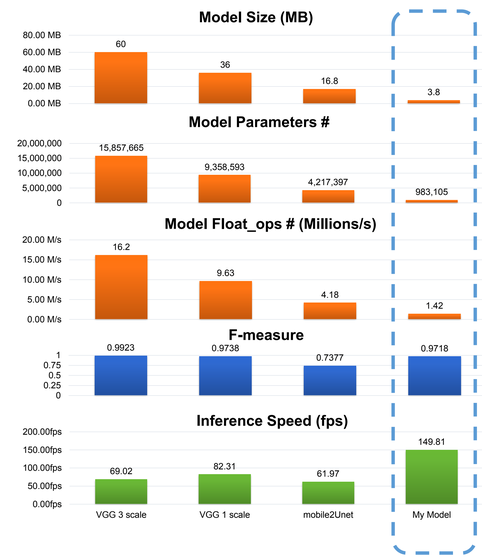
	Training Scheme	F-Measure	Parameters #	Flops(millions/s)	Model Size (bytes)	Inference Speed (fps)
FgSegNet 3-scale	single-scene, RGB	0.9888	15,857,665	16.20	60MB	69.02
FgSegNet 1-scale	single-scene, RGB	0.9738	9,358,593	9.63	36MB	82.31
MobileNet+UNet	single-scene, RGB	0.7377	4,217,397	4.18	16.8MB	57.96
Proposed Model	single-scene, RGB	**0.9718**	**983,105**	**1.42**	**3.8MB**	**149.81**



Our proposed model achieves:

- ✓ **10 times smaller** in Model size and Parameters
- ✓ **10 times reduction** in Floating-point operations (FLOPS)
- ✓ **competitively accurate** in F-measure
- ✓ **2 times faster** in Inference Speed

Comparison in:
- Model size (MB)
- Parameters #
- FLOPS (millions/sec)
- F-measure (0~1)
- Inference speed (fps)



Conclusions and future work

- ✓ **We devise an efficient deep network model** based on depth-wise and point-wise convolution and deconvolution for moving object detection in surveillance video.
- ✓ Experimental studies demonstrated the effectiveness of our proposed model in **smaller model size, less computations, faster speed, competitive accuracy.**
- ➢ For future study, we intend to incorporate **video temporal information** and to **achieve better performance in real-time mobile devices.**

References

[1] Yu Cheng, Duo Wang, Pan Zhou, and Tao Zhang, "A survey of model compression and acceleration for deep neural networks," arXiv:1710.09282, 2017.

[2] Y. Wang, P. M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar, "CDnet 2014: An expanded change detection benchmark dataset," IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work., pp. 393–400, 2014.

[3] Y. Zheng and L. Fan, "Moving object detection based on running average background and temporal difference," Proc. 2010 IEEE Int. Conf. Intell. Syst. Knowl. Eng. ISKE 2010, pp. 270–272, 2010.

[4] Q. Zhou and J. K. Aggarwal, "Tracking and classifying moving objects using single or multiple cameras," Handb. Pattern Recognit. Comput. Vision, 3rd Ed., vol. 0012, pp. 499–524, 2005.

[5] Liu Y and Pados D. A, "Compressed-sensed-domain l1-pca video surveillance," IEEE Trans. Multimedia, 18(3), 351–363 (2016).

[6] Pierantozzi M, Liu Y, Pados D. A, and Colonnese S, "Video background tracking and foreground extraction via l1-subspace updates," Proc. SPIE Commercial + Scientific Sensing and Imaging, Compressive Sensing V: From Diverse Modalities to Big Data Analytics 9857, 985708 (2016).

The paper is accepted by IEEE International Symposium on Circuits and Systems (ISCAS) 2020.

EAPS: Edge-Assisted Predictive Scheduling in WiFi-based IoT Networks

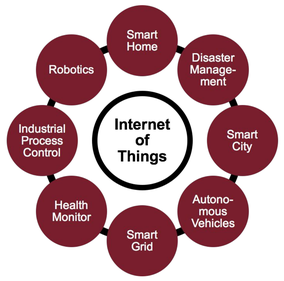
Jaykumar Sheth, Cyrus Miremadi, and Behnam Dezfouli
Internet of Things Research Lab (*SIOTLAB*), Santa Clara University

Santa Clara University
School of Engineering

Motivation

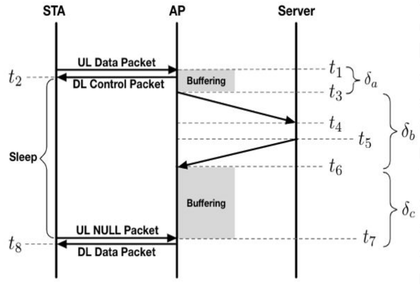
- There will be 50 billion Internet-connected devices by 2020 [1]
- Technology has enabled the production of low-power WiFi stack
- WiFi access points have been widely deployed [2]

LTE/ Cellular	Very expensive, licensed, not energy efficient
Zigbee/ Bluetooth	Low data rate, requires additional infrastructure
WiFi	**• Practically ubiquitous • No additional infrastructure • High data rate • Newly manufactured 802.11 chips are low-power**



- We focus on mission-critical IoT applications
- We are interested in using existing WiFi infrastructures for applications such as medical monitoring and industrial process control

Problem Statement



This Figure displays the end-to-end delay components between a station (STA) and a server including δa (uplink buffering delay), δb (latency over the course of the wired part of the network), and δc (latency at the juncture between the wired and wireless part of the network, a.k.a Access Point).
The prediction of δc is particularly challenging because it is affected by several factors such as traffic rate, channel utilization, and buffering mechanisms employed by the Linux kernel's network layer and wireless NIC driver.

- Once a request is sent by the IoT device, it will have no idea when the response will be sent due to dynamic channel conditions.
- Our study shows that existing 802.11-based IoT devices suffer from high energy consumption, packet loss, and communication delay, especially in dense environments, due to interference and concurrent transmissions.
- Higher amount of traffic due to larger number of WiFi devices → IoT devices suffer from high energy consumption and communication delay

***Objective*: Employ delay prediction and communicate it to IoT device such that it can optimize its sleep/wake schedule to improve the energy-efficiency and timeliness of IoT devices.**

Methodology

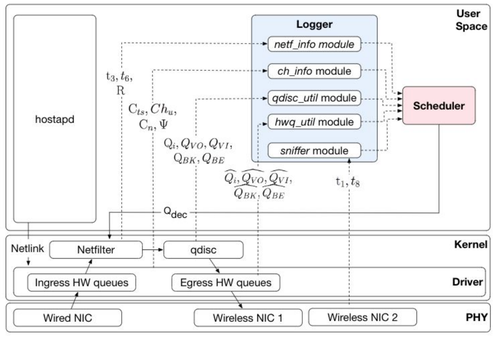
Statistics Collection
- Channel condition information
- Network queue and transmission queue statistics

Delay Prediction
- Statistical learning algorithms
- Neural networks and Long Short-Term Memory algorithms

Inform IoT Device
- Accelerated control packet transmission using a high priority queue

Sleep/Awake Schedule
- The device can transition to low-power mode until the packet is ready to be transmitted to the IoT device

Architecture of Modified Access Point



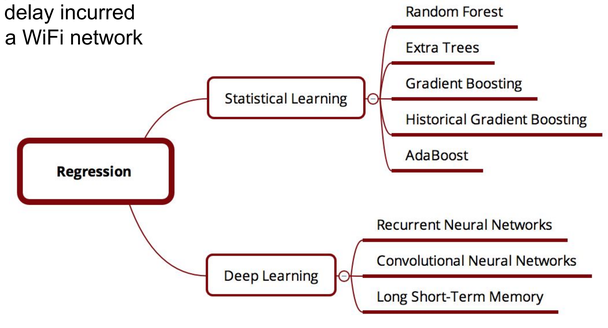
The software architecture of the Access Point.

The Logger communicates with various kernel and user-space components to collect timing information

The Scheduler estimates the sleep duration and conveys it to the STA

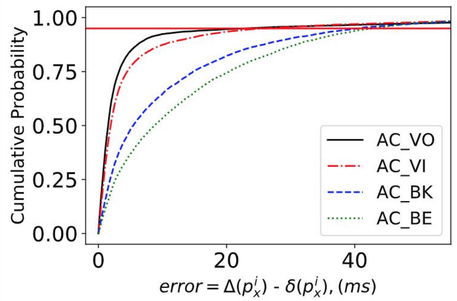
Delay Prediction using Machine Learning Algorithms

The machine learning algorithms employed in order to predict the delay incurred by a packet in a WiFi network

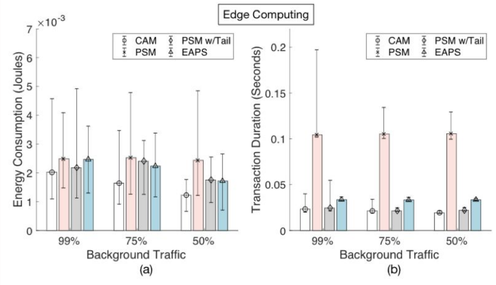


Results

- We have evaluated the optimization provided by EAPS in terms of delay and energy efficiency, compared to state-of-the-art power-savings mechanisms.



The cumulative histogram of prediction errors (predicted delay - actual delay) with respect to four access categories, when using a LSTM algorithm.



Per-transaction energy consumption (a) and delay (b) of EAPS versus CAM, PSM, and PSM w/Tail when the IoT STA is communicating with an edge server
The delay of wired-to-wireless switching is considerably higher than that of AP-server delay. In addition to achieving higher energy efficiency, the transaction delay of EAPS is considerably lower than that of PSM

Conclusion

- We have modified an off-the-shelf access point to predict packet delivery delay and enable IoT devices to perform fine-grained scheduling of their sleep/awake cycles.

Acknowledgement:
This research has been partially supported by the Latimer Energy Lab.

References

[1] Evans, Dave. "The Internet of Things: How the next evolution of the internet is changing everything. "*CISCO white paper,* 1.2011 (2011): 1-11.
[2] Tozlu, Serbulent, et al. "Wi-Fi enabled sensors for internet of things: A practical approach." *IEEE Communications Magazine* 50.6 (2012)
[3] Sheth, Jaykumar, and Behnam Dezfouli. "Enhancing the Energy-Efficiency and Timeliness of IoT Communication in WiFi Networks." *IEEE Internet of Things Journal* 6.5 (2019): 9085-9097

AN IMAGE COMPRESSION SCHEME USING OPTIMIZED BIT PLANE SLICING

Cihan Ruan[1], Meng Wang[2], Nam Ling[3]

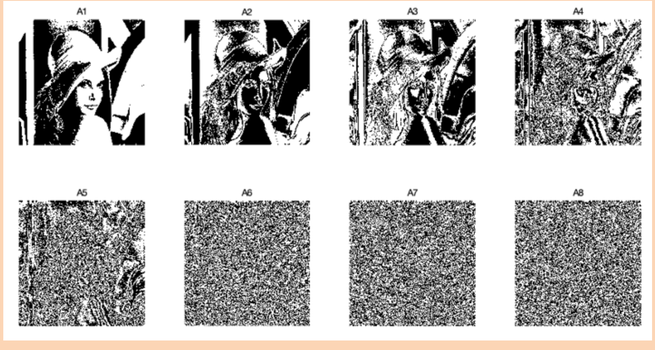
Department of Computer Science and Engineering, Santa Clara University[1,3], School of Medicine, Stanford University[2]

Abstract

- A strategy combining the bit-plane compression and parts of JPEG-2000 to improve the complexity and flexibility compatible with JPEG-2000 at high compression ratios is proposed.
- Our scheme reuses the existing JPEG-2000 encoders to generate bitstream and feature. Two components in addition to JPEG-2000:
  1. an optimization strategy to achieve better coding performance
  2. a novel bit-plane compression method based on bit plane slicing on the whole image

Introduction

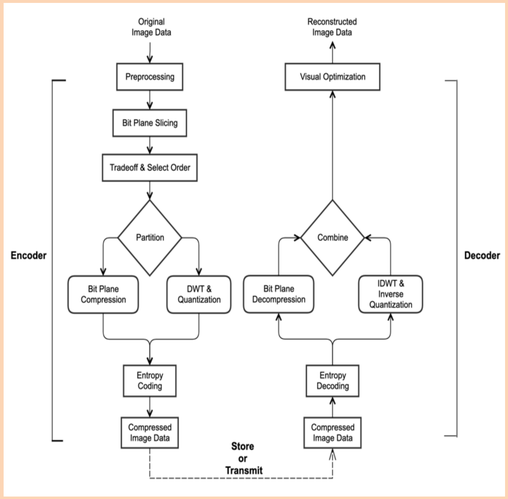
- Nowadays, image compression plays a significant role in both our daily life and lots of research areas. However, the block DCT scheme applied in JPEG will cause boundary blur, and the compression rate of JPEG is not good enough. JPEG-2000 is one of the advantage schemes that could make up for the shortcomings of JPEG. However, the use of DWT in JPEG-2000 increases the coding performance very well and causes high computation complexity, which is a heavy burden at the same time. Moreover, detail information is lost due to the theory of DWT.
- Bit plane slicing is a method of representing an image with one or more bits of the byte used for each pixel. It reduces the original gray level to a series of binary images.



- To balance the visual quality and compression ratio, we designed an optimization strategy to balance the image quality and compression ratio by using bit plane compression theory as:

$$\mathrm{A}=\mathrm{A}^{(0)}+\mathrm{R}(\mathrm{A})^{(0)} \qquad A = a_1A_1 + a_2A_2 + \cdots + a_nA_n + R_n$$

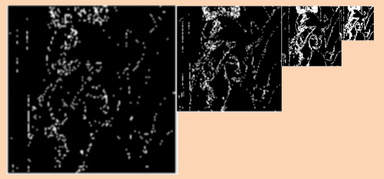
- We adopted the idea of divide and conquer: most of the information is stored in the top 4 bit plane images. To decrease the complexity of a whole image coding, we decided to partition it into 2 parts: 1. the compression of top N layers bit-plane images; 2. the DWT compression of the residual.
- DWT is a good tool for dealing with residual. And Daubechies 9/7 is selected.



- And for the bit-plane compression process, we use inverse Kronecker product to separate the boundaries and derivatives .

$$\begin{cases} A=\partial A+\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}\otimes P(A) \\ P(A)=\partial^2 A+\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}\otimes P^2(A) \\ P^2(A)=\partial^3 A+\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}\otimes P^3(A) \\ \cdots \end{cases}$$

$$\hat{A}=\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}\otimes\left\{\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}\otimes\left[\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}\otimes P^3(A)+\widehat{\partial^3 A}\right]+\widehat{\partial^2 A}\right\}+\widehat{\partial A}$$

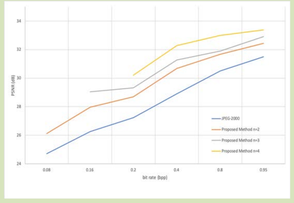


Results

- For an original image A, let $A_{1/64}$ be the compressed image of A through using the D9/7 associated to JPEG-2000.
- To reduce complexity, we use spectrum expansion method (SEM) to replace D9/7.
- Compared to JPEG-2000, our proposed method shows the improvement at high compression ratio.

JPEG-2000		Proposed Method					
Bitrate	PSNR	n=2		n=3		n=4	
		PSNR	Gain	PSNR	Gain	PSNR	Gain
0.08	24.86	25.46	0.60	25.96	1.10	-	-
0.20	28.95	29.82	0.87	30.17	1.22	30.23	1.28
0.40	31.44	32.16	0.72	32.73	1.29	32.46	1.02
0.80	33.21	33.89	0.68	34.08	0.87	34.07	0.86
Average	-	-	0.72	-	1.12	-	1.05

Average results on the test images with different number of n



Rate-Distortion curves of JPEG-2000 and the proposed method (Goldhill)

Conclusion

- A novel lossy image compression scheme is introduced which divides an image into 3 bit planes MSBs plus a residual, then we deal with the bit-planes part and the residual respectively.
- The compression rate and the quality of result is good.
- For now, our theory has only been applied on images. In the future, we would like to expand it into the video coding area.

Sparse Coding of Intra Luma Prediction Residuals for Screen Content Coding

Michael Schimpf; Advisor Prof. Dr. Nam Ling
Department of Computer Science and Engineering

Santa Clara University
School of Engineering

Abstract

- In high efficiency video coding (HEVC), intra frames are divided into blocks, these blocks are then predicted and subtracted from the original, which results in a residual block, which is then transformed by the HEVC core transforms, quantized, and entropy coded [1].
- The quantization process typically reduces high frequency components, which results in high compression, but adds more blurriness and blockiness.
- HEVC screen content coding (HEVC-SCC) is an extension to HEVC which adds sophisticated compressing method for computer generated content [2] which has large uniform areas, but also many areas with higher frequencies because of sharp/complicated edges and text.
- Sparse coding uses a dictionary to transform a block into the sparse domain [3] and can usually better preserve high frequency components.

Sparse Coding

#367

	1	2	3	4	5	6	7	8
1	0	0	80	100	87	17	50	10
2	0	1	55	50	7	5	0	0
3	0	0	30	127	70	2	0	0
4	0	0	2	20	50	0	0	0
5	0	0	30	125	50	0	0	0
6	0	0	20	50	10	0	0	0
7	0	0	5	25	0	0	0	0
8	0	0	0	0	0	0	0	0

#899

	1	2	3	4	5	6	7	8
1	0	0	40	27	40	117	50	80
2	0	2	55	77	120	9	1	0
3	0	0	70	0	20	8	0	0
4	0	0	88	107	30	0	0	0
5	0	0	50	2	29	0	0	0
6	0	0	30	60	30	0	0	0
7	0	0	30	15	0	0	0	0
8	0	0	0	0	0	0	0	0

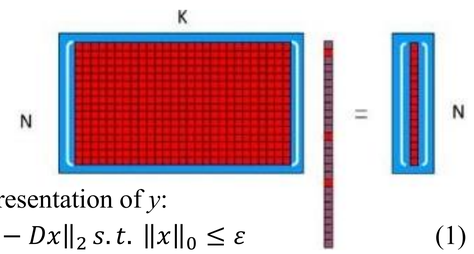
	1	2	3	4	5	6	7	8
1	0	0	120	127	127	127	100	90
2	0	3	110	127	127	45	1	0
3	0	0	100	127	90	10	0	0
4	0	0	90	127	80	0	0	0
5	0	0	80	127	79	0	0	0
6	0	0	50	110	40	0	0	0
7	0	0	35	40	0	0	0	0
8	0	0	0	0	0	0	0	0

An 8x8 block can be represented by two indices i.e. 367 and 899 plus a coefficient per index
2 indices + 2 coefficients: (2*10bits) + (2 * 8bits) = 36 bits
64 8-bit values: (64 * 8bits) = 512 bits

Sparse Coding

- Sparse coding is a process of calculating atoms for sparse vector x of the signal y with a dictionary D and uses as few atoms as possible; the problem is NP-hard [5].
- An atom contains the index to the dictionary entry as also the coefficient which represents the magnitude of which the dictionary entry is later linearly added to reconstruct the signal [6].
- The greedy strategy called orthogonal matching pursuit (OMP) is based on Eqs. (1) and (2). With every iteration of OMP, one dictionary atom is chosen which represents signal y closest, then this finding is subtracted from current signal y and the process continues with the next iteration and the next best candidate to fit the new residual as best as possible.
- With every iteration of adding a new atom, the error decreases until it reaches below a predefined error ε.
- Signal y can be reconstructed as a linear combination of the atoms of sparse vector x and dictionary D.

$y \in \mathbb{R}^n$ — *y = signal, N = signal dimension*
$D \in \mathbb{R}^{n\,x\,K}$ — *D = dictionary, K = # of atoms*
$x \in \mathbb{R}^K$ — *x = vector of indices/coefficients*



Approximation representation of y:

$$\min_x \|y - Dx\|_2 \; s.t. \; \|x\|_0 \leq \varepsilon \quad (1)$$

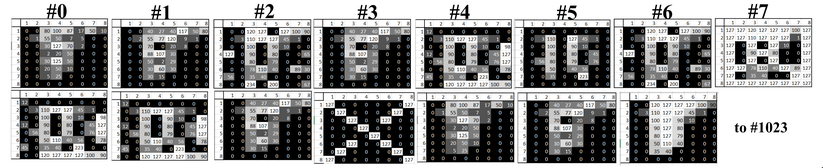
$$\min_x \|x\|_0 \; s.t. \; \|y - Dx\|_2 \leq \varepsilon \quad (2)$$

$\|.\|_2$ represents the l_2 norm and $\|.\|_0$ the l_0 pseudo norm.

Proposed Method

- The proposed method applies sparse coding in place of 8x8 intra luma predicted 8x8 HEVC core transformed residuals adaptively by integrating OMP into HEVC-SCC.
- Additionally a category approach to the training and test process is adopted. These CTC [4] categories are animation, mixed content, text, and graphics with motion.
- The selection process between HEVC core transforms and OMP is based on a standard RDO decision. RDO delivers an optimal approximation by minimizing R-D cost J(k): J(k) = D(k) + ⅄R(k) where R(k) is the number of bits, D(k) is the mean square error (MSE) between original and reconstructed signals, and ⅄ represents the Lagrangian multiplier.
- The OMP algorithm computes, after each iteration, an HEVC/HEVC-SCC RDO cost calculation and compares it with the cost of the initial executed HEVC core transform.
- In general, for the first OMP iteration, cost is high because the distortion is high, but it decreases with every iteration and every atom, until it reaches an optimal point.
- When it reaches an optimal point and it is lower than the HEVC core transform cost, the algorithm stops and OMP instead is encoded into the bitstream.

Dictionary



Experimental Results

	All Intra				
	G/Y	B/U	R/V	Enc Time[%]	Dec Time[%]
YUV, text & graphics with motion, 1080p & 720p	-0.4%	-0.2%	-0.3%	162%	118%
YUV, mixed content, 1440p & 1080p	-1.3%	-0.7%	-0.8%	160%	107%
YUV, Animation, 720p & 768p	-1.6%	-0.8%	-0.8%	162%	126%
Overall	-0.8%	-0.4%	-0.5%	161%	117%

ompFlag1 (8x8)	ompFlag0 (8x8)	Sum 100%	% OMP	% Rest		
5,790,061	19,568,583	25,358,644	22.83	77.17	100.00	ALL
1,485,506	6,595,375	8,080,881	5.86	26.01	31.87	22
2,065,498	5,058,954	7,124,452	8.15	19.95	28.09	27
1,135,026	4,475,345	5,610,371	4.48	17.65	22.12	32
1,104,031	3,438,909	4,542,940	4.35	13.56	17.91	37
5,790,061	19,568,583	25,358,644	22.83	77.17	100.00	Sum

Conclusion & Future Work

- The results show that sparse coding of 8×8 intra luma predicted residual blocks improves BD-rate saving well for the A and MC category, but not for the TGM category because the IBC and PM extension already efficiently compressed text well.
- The OMP method increases execution time because of the higher complexity.
- Entropy coding needs to be optimized and the proposed method will be extended to 4x4, 16x16, 32x32, and for chroma.

References

[1] Gary J. Sullivan, Jens-Rainer Ohm, Woo-Jin Han, and Thomas Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," IEEE Transactions on Circuits and Systems for Video Technology, Volume 22, Number 12, December 2012.
[2] Jizheng Xu, Rajan Joshi, and Robert A. Cohen, "Overview of the Emerging HEVC Screen Content Coding Extension", IEEE Transactions on Circuits and Systems for Video Technology, Volume 26, Issue 1, Jan. 2016.
[3] M. Aharon, "Overcomplete Dictionaries for Sparse Representation of Signals," in PhD thesis, The Technion Israel Institute of Technology, 2006.
[4] Haoping Yu, "Common test conditions for screen content coding", JCTVC-Z1015, 2017.
[5] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," Journal of Constructive Approximations, Volume 13, pp. 57-98, 1997.
[6] M. Aharon, M. Elad and A. M. Bruckstein, "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation," IEEE Transactions on Signal Processing, Volume 54, Issue 11, pp. 4311–4322, November 2006.
[7] HEVC Reference Software HM-16.18+SCM-8.7 (2018 January) [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.18+SCM-8.7

Channel-Separable Neural Network for Block-based Image Compression

Pengli Du, Dr. Ying Liu
Department of Computer Science and Engineering
Santa Clara University
School of Engineering

Background

- The main frame of image compression is based on the autoencoder. The schema is as follows.
- The image is fed into the encoder, and the decoder reconstructs the image. The key components are the neural network structures of the encoder and decoder.

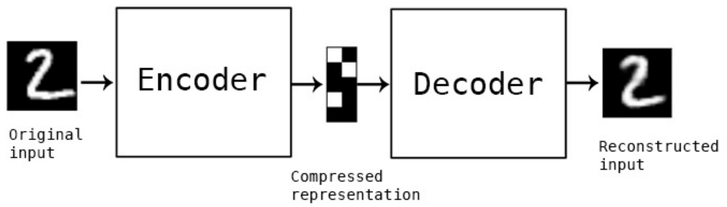


Fig 1. Autoencoder Schema

Objectives

- Use fully connected dense layers to work as the encoder, which is a trainable encoder. The deterministic encoder is also tested. Only one kind of decoder is used, which is a trainable decoder.
- In addition, split color image into RGB channels and compress them separately. Then concatenate these three channel images into one to get the reconstructed image. It's compared with the reconstructed result of the color image without being split.

Research Design

- According to the choice of encoder and whether to split the color image or not, 4 methods are proposed.

Network Structure

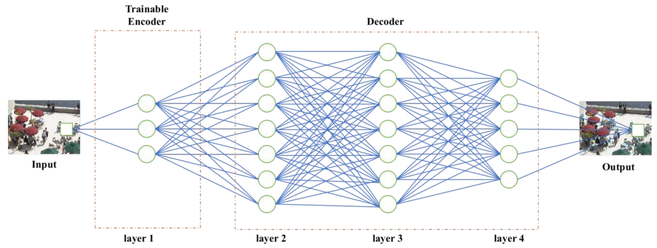


Fig 2. RGB_trainable_encoder_decoder

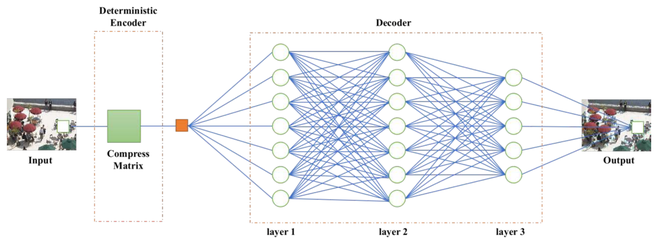


Fig 3. RGB_deterministic_encoder_decoder

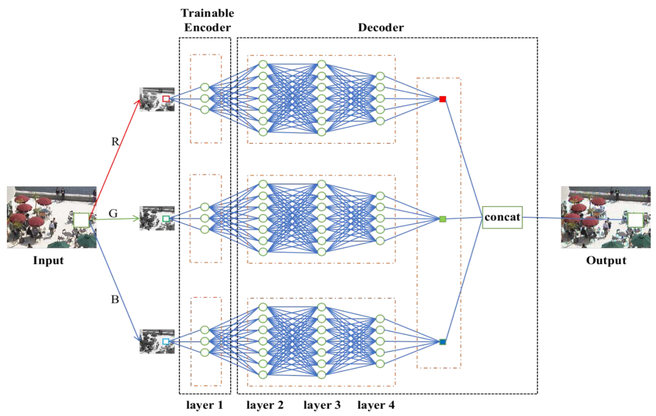


Fig 4. split_RGB_trainable_encoder_decoder

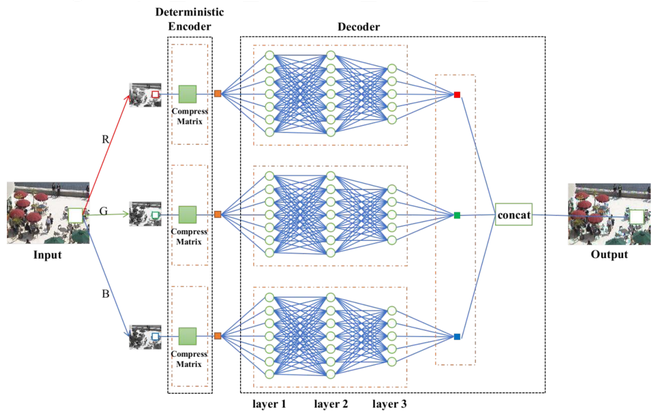


Fig 5. split_RGB_deterministic_encoder_decoder

Experiments Results

Table 1. The average PSNR value of the testing data

Method	0.01	0.02	0.0625	0.125	0.25	0.375	0.5	0.6
sRGB_tED	31.501	34.776	40.832	44.868	52.541	74.026	72.924	73.725
RGB_tED	28.512	32.500	38.706	42.020	46.581	50.374	62.699	63.202
sRGB_dED	22.015	22.909	23.981	24.354	24.558	24.828	24.763	24.922
RGB_dED	21.251	21.526	22.537	22.863	23.329	23.519	23.768	23.707

Table 2. The average SSIM value of the testing data

SSIM CR	0.01	0.02	0.0625	0.125	0.25	0.375	0.5	0.6
sRGB_tED	0.857	0.924	0.977	0.990	0.998	0.999	0.999	0.999
RGB_tED	0.721	0.878	0.967	0.984	0.994	0.998	0.999	0.999
sRGB_dED	0.185	0.263	0.370	0.402	0.419	0.439	0.435	0.448
RGB_dED	0.074	0.088	0.210	0.246	0.299	0.321	0.349	0.341

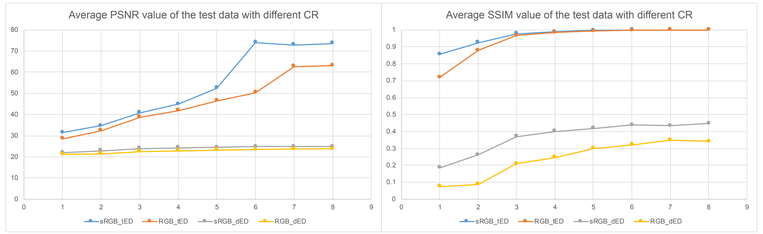


Fig 6. The result of PSNR and SSIM value of the model with 4 methods

Conclusions & Plan

- Trainable encoder is better than deterministic encoder and splitting the color image into 3 channels and compressing them separately does help the reconstructed result.

Fig 7. Result of the best two methods, (**sRGB_tED** and **RGB_tED**) with different compression ratio, which are 0.01, 0.02, 0.0625, 0.125, 0.25, 0.375, 0.5 and 0.6.(From the left side to the right)

- In the future, improve the network structures and apply image compression techniques into videos to recover the consecutive frames.

A Novel Deep-Learning Approach To Encode Location Information And Its Application On Bike Share Service

Yuan Wang[1]*, Yinan Ling[2]*, Chenwei Wang[3]*, Hsin-Tai Wu[3], Yokoyama Keita[3], Yi Fang[1]

[1]Dept. of Computer Science and Engineering, Santa Clara University

[2]Data Science Institute, Columbia University

[3]Docomo Innovations, Inc.

Santa Clara University
School of Engineering

Abstract

- We developed a universal location encoder capable of generating efficient location representation via numerical vectors.
- We proved that the location-based vector representation carries both spatial geometry and semantic information.
- We applied them to a downstream task of seeking new locations for NTT DOCOMO's share-bike business.

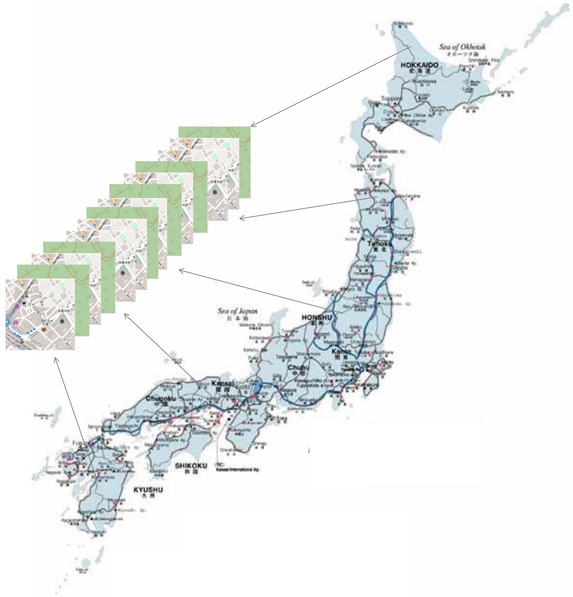
Introduction

Since the locations are usually represented by a pair of longitude and latitude which does not carry any information, we aim at understanding the spatial and semantic information of the locations.

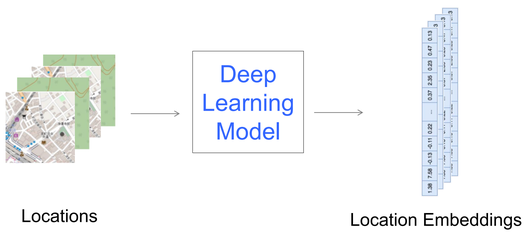
latitude/longitude
(35.68, 139.76)

Methodology

We collected a dataset from a Japanese map, including the map tile images, the point of interest information and the map metadata.

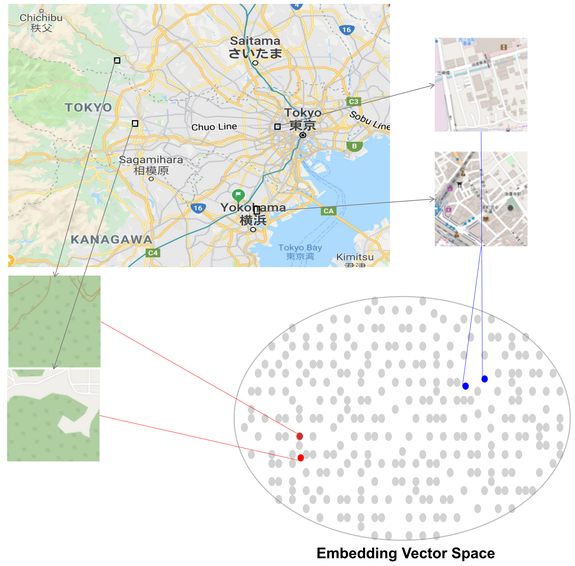


Then we trained a multi-label classification model using Resnet-50 using the dataset.

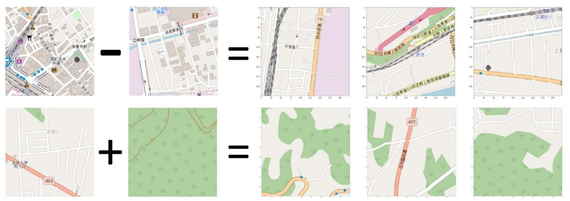


Evaluation of the Embeddings

- Similar map tile images tend to stay closer in the embedding vector space, regardless of whether their geo-locations are closer or not.

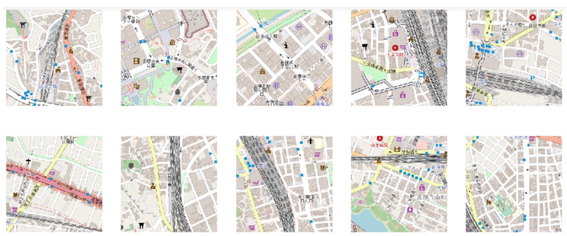


Location Embeddings Carry Semantic Information:

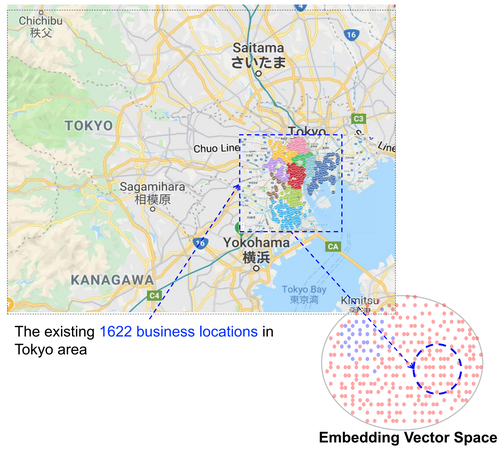


Application to Bike Share Service

Recommendation of the New Service Locations

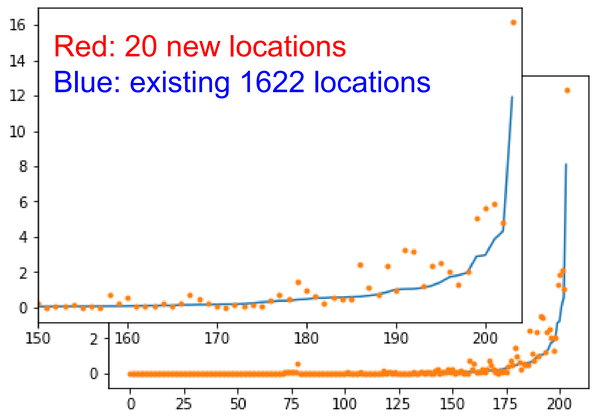


Finding new locations is equivalent to identifying the candidates closest to the existing locations in the embedding vector space.



The existing 1622 business locations in Tokyo area

Evaluation of the Recommendation



The existing 1622 and the 20 new recommended locations have the similar POI count 204-dimensional vector distribution.

Future Plans

The use of the location embedding still has a lot of potential. In the future, we would perform more experiments on other location-related tasks.

*Authors contributed equally to this research.

Multi-label Classification for Short Text Utilizing the Label Correlations

Zhiyuan Peng[1], Yi Fang[1]
[1]Dept. of Computer Science and Engineering, Santa Clara University

Santa Clara University
School of Engineering

Abstract

Tagging the shot text, such as tweet and message, with multi labels is challenging due to less sequence information provided by the short text. This paper presents a novel method using a tree structure to represent the relationships among labels and the max-product algorithm for inference.

Introduction

Unlike long texts, short texts could provide less sequence information, some of which are even unstructured data and the order of which has no influence on the meaning of the texts. Take the following two searching queries as an example:

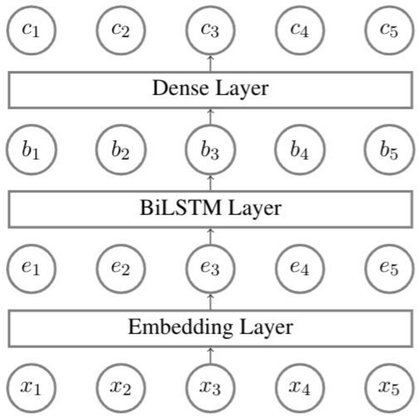
- Sony TV HD 60 inch
- HD TV Sony 60 inch

We want the search engine to return the same result: TV with brand Sony, resolution HD and size 60 inch no matter which one of the two queries we use.

For multi-label classification, the labels are not all independent. We can use the correlations among labels to improve the performance of multi-label classification for short text. We summarize our main contribution as follows:

- We use the maximum spanning tree algorithm to generate a tree structure to represent the relationships between labels.
- For each edge in the tree, we train a binary classifier and use the max-product algorithm for inference.
- Experiments results on three short text datasets show that the correlations between labels do help improve the accuracy of the multi-label classifier.

Widely used Model



Embedding Layer: Embedding layer maps each token into a vector representation which keeps the semantic similarity. Here, we use the glove.6b.100d.txt to map each word to a vector.

LSTM Layer: Long Short Term Memory (LSTM) is a kind of Recurrent Neural Network (RNN) which is used to capture the long-term dependencies and avoid the problems of RNN. The chain structure of LSTM is naturally suited to capturing the sequence information in the text. After the LSTM layer, we get a vector which is learned from the text

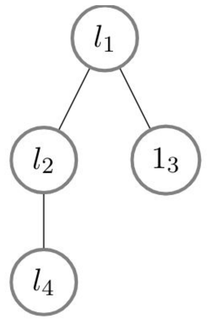
Dense Layer: Dense Layer is a fully connected one layer neural network. We use Sigmoid activation function. The number of the outputs is equal to the number of the classes.

Proposed Model

Tree Structure: For each pair of labels, we calculate the co-occurrence score:

$$Score\,(l_i, l_j) = \frac{count(l_i, l_j)}{min(count(l_i), count(l_j))}$$

After we use the Maximum Spanning Tree algorithm, we can get a tree structure like this:



Then, we convert the tree to a factor tree.

Binary Classifier: The structure of the binary classifier for edge is almost the same as that of the original model. The only difference is the dimension of the output. Here, for the binary classifier, there is only one output.

Inference: We use the Max-Product algorithm to get the most possible label sequence. The message from node i to the factor a is like this:

$$\mu_{i\to a}(\mathbf{y}_i) = \psi(\mathbf{y}_i) \prod_{c\in N(i)\backslash a} \mu_{c\to i}(\mathbf{y}_i)$$

The message from factor a to node j can be computed as:

$$\mu_{a\to j}(\mathbf{y}_j) = \max_{\mathbf{y}_i} \left(\psi_a(\mathbf{y}_i, \mathbf{y}_j)\mu_{i\to a}(\mathbf{y}_i)\right)$$

At each max operation, we keep the value that achieves the max value.

At the root we multiply all the incoming messages and get the maximum possibility. To get the most likely labels, we can go back to find the labels.

Result

We test our method on three short texts and the description of the datasets is:

Name	Description
Movie comment	16225 comments 6 labels
SemEval-2018	10986 tweets 11 labels
NTCIR-13	1,920 tweets 8 labels

We also use the true edge information for inference to get the upper bound of our method. The comparison of the accuracy is as follows:

Name	Original Method	Our Method	Upper Bound
Movie comment	0.822	0.843	0.911
SemEval-2018	0.778	0.825	0.903
NTCIR-13	0.834	0.872	0.953

Conclusion

Compared with the widely used model, our method exploring the correlations among labels improves the performance of the multi-label classifier. The expense is that we need to train a binary classifier for each edge in the tree.

Modeling the Dynamics of Personal Expertise

Archana Godavarthy, Xuyang Wu, Yi Fang
Department of Computer Science and Engineering, Santa Clara University

Santa Clara University
School of Engineering

Background

Personal Expertise

- Often evolves over time
- Capturing how personal expertise evolves can better characterize expert profiles

Factors can affect the dynamics of personal expertise

- **Similarity:** The similarity between prior personal expertise and the new interested areas
- **Popularity:** The popularity of different personal expertise by different time
- **Personality:** The personal efforts to explore new areas and skills
- **Co-partnership:** The colleagues or partner efforts which influence of the personal expertise

Expertise Estimation Model

Model:

- A probabilistic model to characterize how people change or stick with their expertise
- Four factors are taken into consideration: personality, similarity, popularity and co-partnership
- Used to predict what are the next expertise areas a given expert will work on, and whether there will be a big change in her expertise areas

Definition:

- a_{t+1} is the expert's expertise areas in year $t+1$, a_t is the expert's expertise areas in year t.
- $P(a_t|e)$ is the probability that expert e's current area in year t is a_t .
- $P(a_{t+1}|a_t, e)$ is the probability that she will work on the area a_{t+1} given her current area a_t .
- $P(a_{t+1}|e)$ is the probability that the expert e will work on the area a_{t+1} in year $t+1$.

$$P(a_{t+1}|e) = \sum_{a_t} P(a_{t+1}|a_t, e)P(a_t|e)$$

Features

Personality:

The expert e chooses the next area a_{t+1} given the current a_t as the following generative process:

- Choose to stay in the current expertise area set A_t with the probability $P(a_{t+1} \in A_t \,|a_t, e)$
  - Choose a specific topical area a_{t+1} from A_t with the probability $P(a_{t+1}|a_t, e, a_{t+1} \in A_t)$
- Choose to change to the new area set $\bar{A}_t$ with the probability $1 - P(a_{t+1} \in A_t \,|a_t, e)$
  - Choose a specific topic area a_{t+1} from $\bar{A}_t$ with the probability $P(a_{t+1}|a_t, e, a_{t+1} \in \bar{A}_t)$

$$\begin{aligned} P(a_{t+1}|a_t, e) &= P(a_{t+1} \in A_t|a_t, e)P(a_{t+1}|a_t, e, a_{t+1} \in A_t) \\ &+ P(a_{t+1} \in \bar{A}_t|a_t, e)P(a_{t+1}|a_t, e, a_{t+1} \in \bar{A}_t) \end{aligned}$$

Similarity:

The expert e chooses a new area, i.e., $a_{t+1} \in \bar{A}_t$, we will consider the dot product similarity between the new area and the current areas:

$$Sim(a_{t+1}, a_t) = a_{t+1} \cdot \sum a_t$$

- a_{t+1} is the new area's vector embedding in year $t+1$, based on pre-training area vector embedding.
- $\sum a_t$ represents the sum of area's vector embedding in year t.
- $Sim(a_{t+1}, a_t)$ measures the similarity between two areas a_{t+1} and a_t.

Popularity

The expertise area gets popular or trendy, people are likely to move to the area regardless of their prior background (e.g., a_t), calculated based on the relative frequency of the area appearing in year t.

$$Pop(a_{t+1}) = \frac{N_{a_{t+1},t}}{N_t}$$

- $Pop(a_{t+1})$ measures the popularity of a_{t+1}.
- $N_{a_{t+1},t}$ is the number of times that a_{t+1} occurs in e's publications in year t.
- N_t is the total number of times that any topical area occurs in e's publications in t.

Co-partnership

The expert e's expertise is influenced by her co-authors. We will consider the prior co-authorship based on publications, aggregated e's co-authors appearing in year t and get the embedding representation.

$$h^k_{N(e)} \leftarrow AGGREGATE(\{h^{k-1}_u, u \in N(e)\})$$

- $h^k_{N(e)}$ is the co-authorship embedding representation of expert e.
- $\{h^{k-1}_u, u \in N(e)\}$ is expert e's immediate neighborhood (co-authors) embedding representation.
- $AGGREGATE()$ is the aggregator functions which aggregate the co-author information (e.g., Mean aggregator, LSTM aggregator, Pooling aggregator).

Experiments

Predictive language model: obtain the predicted probability for e over topical areas a_{t+1} (in year $t+1$); based on it, we can further estimate e's probability over words in $t+1$

Logistic Regression: change the predict model to logistic regression, use the topic areas a_{t+1} (in year $t+1$) as the label to predict

Deep Learning Algorithm: change the predict model to 3-layer deep learning model, use the ReLu function as activation function

Dataset: 500 researchers on "information retrieval" with their published papers

Metrics: KL divergence

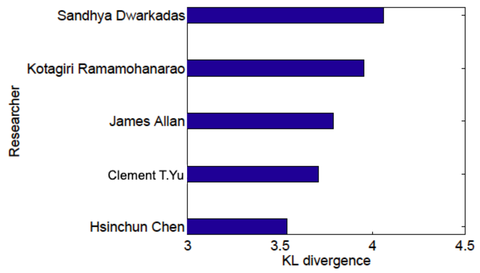
Learning Models: Predictive language model & Logistic Regression

Results

The 5 researchers with the largest predicted KL divergence from 2012 and 2013

Researcher	Keywords in 2012	Predicted Keywords in 2013
Dawei Song	Concept hierarchy Log analysis Query suggestions	Query expansion Search log mining Query suggestions
Yoelle S. Maarek	Community QA Web retrieval Query analysis	Community QA User interaction Query analysis
Clement T. Yu	Coreference resolution Topic model Diversionary comments	Blog retrieval Faceted blog distillation Subjectivity

Top areas in 2012 and predicted top areas in 2013 for three researchers



Improvements and Future Plans

- Explore more factors in modeling expertise dynamics.
- Conduct a more comprehensive set of experiments to evaluate the proposed model on large-scale testbeds.

Hydration Automation

Peter Ferguson, Zach Cameron, Nicholas Kniveton, Will Tuttle, Jesse Mayer
and Prof. Navid Shaghaghi

Problem

- Agricultural systems are dependent on water.
- These systems are equipped with irrigation systems that store water and manage the distribution.
- Modern irrigation systems are equipped with many sensors and actuators to manage and monitor water.
- Industrial-level systems on the market are expensive, complex, energy inefficient, and are often manually operated.
- Systems are inaccessible to smaller farms or ranches.

Figure 1: Pastures

Figure 2: Farmlands

Solution

- We have designed and are building a low cost, small factor, smart, and sustainable system called Hydration Automation (HA) to monitor and manage water level and distribution in water tanks.

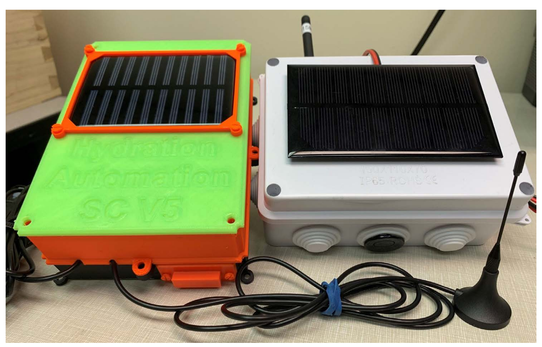


Figure 3: 3D printed device (left) and prefabricated prototype device (right)

System Overview

Components

- Sensing Units (SUs)
- Relay Units (RUs)
- Actuating Unit (AU)

We designed each unit in the system to be modular and have easily replaceable components. The SUs are attached to water tanks to measure water levels. The data is sent over the wireless subsystem to the AU. The RUs are used as relay nodes in the system to transport the water level data over long distances or around large objects. The AU collects analyzes the data from all the tank sensors and controls valves or pumps to distribute water to the different locations in the system. The system is energy efficient and the simple Arduino base allows for easy addition and removal of units to the system.

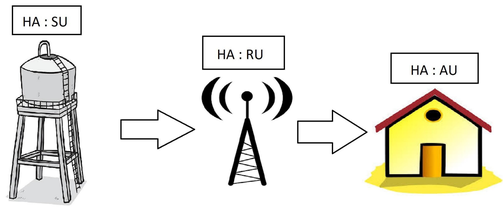


Figure 4: System overview diagram

Software

SU Software

The SUs spend most of their time in a low power (sleep) mode and only awaken periodically for measurements and transmission to conserve power. The SU uses a real-time clock (RTC) attached to an interrupt to wake up the board from sleep mode.

RU Software

The RU follows a repeater module. It will repeat the messages it receives and send the message to the nodes that are not in its subnet.

Communication Protocol

ÂB is an Energy Aware Communication Protocol (EACP) which has been developed for the communication subsystem of HA. It runs atop the LoRa transport layer and a wireless physical layer utilizing RFM95 Radio Modules in the 915/916 MHz band. The subsystem protocol reduces memory consumption and is power efficient allowing for its use in IoT applications.

ÂB's trademark feature is its ability to takes advantage of each individual network component's low power sleep cycle using sleep cycle duration synchronization without a synchronized global clock. It accomplishes this by performing real time calculations for how long each node in the network can sleep and then communicating specific instructions to each node on the network.

Since ÂB can send RUs to sleep as well, they will only be monitoring packets when they are awake. And since the wake periods can be algorithmically shortened to be very small, more energy can be saved there. Furthermore, due to the shortness of the wake periods, the AU is able to create a sleep schedule for its SUs and RUs that results in no conflicts. Therefore, the times where an RU may hear a transmission that was not originally intended for it are further minimized.

Destination Address (variable size)	'\0' (null character, 8 bit)	Source Address (variable size)	'\0' (null character, 8 bit)	TTL (16 bits)	Message Type (8 bits)	Payload (variable size)	'\0' (null character, 8 bit)

Figure 6: Packet Layout

Packet Design

The packets are designed to work with the variable-sized address that our network has. The null character is used to show when variable length payloads are done. The destination address is the receiving device's address in the network; this is assigned by the AU. The source address is the sender's address in the network; it is also assigned by the AU. One of those two will always be the AU which has a unique address. The TTL is the time to live; it says how many hops the packet has left until it gets to the destination. It calculates TTL through the SU's address length (SU's address is a string). The message type tells the receiver what type of data is being sent. The final part of the packet is the payload which is the data being sent, followed by a null character to let the receiver know that the payload is done.

Hardware

SU Hardware

- Whisper Node with LoRa
- JSN-SR04T ultrasonic sensor
- Lithium Battery
- 0.5 watt Solar Panel

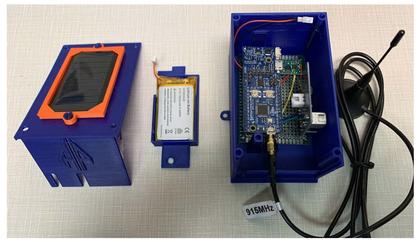


Figure 6: Opened SU device

RU Hardware

- Whipser Node with LoRa
- Lithium Battery
- 0.3 watt Solar Panel

Casing

- 3D Printed
- Weather Proof
- Recyclable
- Small Scale

Figure 7: Opened RU device

Figure 8: RU case design

AU Hardware

- Whisper Node with LoRa
- Valve Controller
- Pump Controller
- Solar Panel

The AU will consist of actuators attached to an electric valve controller which will be powered by the same LoRa-based board as the other components, with a larger lithium battery solar panel. The pump controllers allow the users to control the distribution of water to the tanks.

Work In Progress

Security

We are planning to add verification to when units are added to the AU. Secondly, we plan to implement a system to track when/how many messages are sent from each SU, this allows for us to know if a message is sent from a valid source preventing spoofing. Lastly, we plan to modify the casing to prevent/alert the user to tampering on the system.

3D Nanocarbon Interconnects

Yu Zheng, Parth Shah, Dongmeng Li, and Cary Y. Yang

Center for Nanostructures, Santa Clara University, Santa Clara, CA

Abstract

The superior thermal and electrical properties of carbon nanotubes (CNTs) and graphene present great promise for these nanocarbons as potential candidates to replace copper (Cu) and tungsten (W) in on-chip interconnects in next-generation integrated circuits. Cu and W face increased reliability challenges in the nanoscale as a result of electromigration failures at high current densities. Due to its superior electrical and mechanical properties as well as much higher current-carrying capacities, CNT has been demonstrated to be a serious contender in on-chip interconnect vias. However, the main challenge to functionalizing CNT and graphene interconnects is the metal-nanocarbon contact resistance. To mitigate such challenge, a seamless three-dimensional all-carbon interconnect structure is conceived and fabricated by growing CNTs directly on multi-layer graphene (MLG). This three-dimensional (3D) structure can potentially yield low resistance due to the strong C-C sp^2 bonding in CNT and graphene, and presumably across the CNT/graphene interface. Our study focuses on improving the CNT/MLG interface by varying the CNT growth conditions, with minimum or no damage to the MLG underlayer, thus reducing the resistance of the test structure. The 3D structure can serve as a building block for an on-chip interconnect network.

Background

- Figure 1 demonstrates that interconnects become the dominant performance-limiting factor for chip technology nodes beyond 10 nm [1].
- One major challenge in sub-10 nm integrated circuit (IC) technology is the reliability and performance problems encountered in current interconnect material Cu (and W), as illustrated in Figure 2 [2,3].
- Nanocarbons, such as CNTs and Gr, are potential replacements due to their superior electrical, mechanical, and thermal properties [2,4-8]. An all-carbon 3D structure consisting of vertical CNTs deposited on graphene would be an appealing candidate.

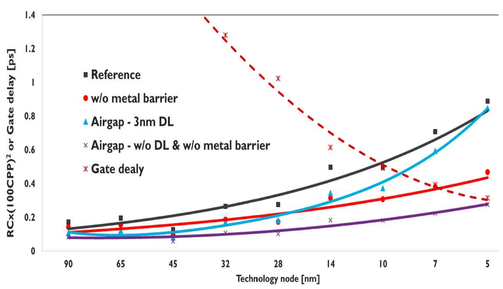


Figure 1. Comparison of delays from various interconnect configurations with gate delay [2]. Reference data pertain to Cu interconnects.

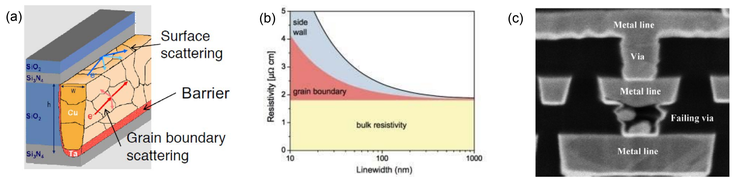


Figure 2. (a) Defects formed in Cu interconnect **(b)** Schematics illustrating increased resistivity in nanoscale Cu interconnect due to enhanced scatterings at surface and grain boundaries [2]. **(c)** Failure of Cu via at high current density [3].

Objective

The objective of this project is to design a fabrication process for CNTs grown directly on MLG that results in improved CNT-underlayer contact from those obtained for CNT vias connecting conventional metals [5].

Test devices are fabricated in TENT Laboratory located at NASA Ames Research Center and characterized at the Center for Nanostructures on the SCU campus.

CNT Vias

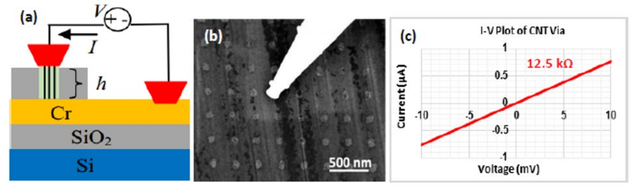


Figure 3. (a) Schematic of I-V measurement setup. (b) SEM image showing nanoprobe in contact with 60 nm CNT via without metallized top Pt contact. (c) Typical I-V plot of a 60 nm via without metallized top Pt contact [7].

CNT via Resistance

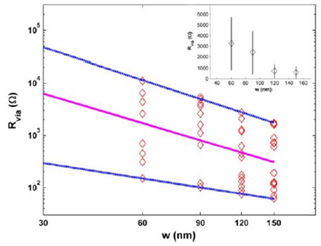


Figure 4. Measured CNT via resistance versus via width (data points) and statistical linear regression fit of the measured data (solid line). The dotted lines show the fits based on the lowest and highest resistances, respectively, for each width. The lowest resistance obtained for a 60 nm via is 150 Ω and the lowest extrapolated resistance for a 30 nm via is 295 Ω. The inset shows the arithmetical mean (data point) and standard deviation (vertical bar) for each via width [5].

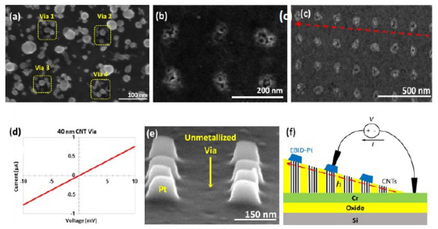


Figure 5. (a) Top-view SEM image of 40 nm vias after CNT growth. (b) CNT vias of 40 nm after Al_2O_3 filling and polishing. (c) Top-view of wedge formed after polishing and before Pt cap deposition. (d) Typical measured I–V behavior that yields via resistance R_{via} before Pt top-contact metallization. (e) Glancing view of CNT vias after Pt top-contact metallization. (f) Cross-sectional schematic of wedge with nanoprobing setup showing varying via height h used for contact resistance extraction [7].

Performance and Reliability of CNT vias

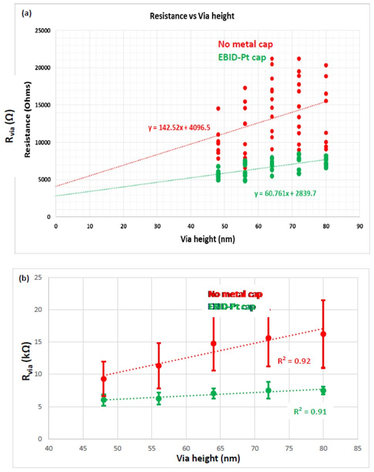


Figure 6. (a) Measured via resistance R_{via} versus via height for 40 nm CNT vias with and without metallized top contacts. Linear fit yields an average contact resistance R_C from the R_{via} intercept. (b) Mean and standard deviation of measured resistances for each via height with $R^2 > 0.9$ for both data sets.

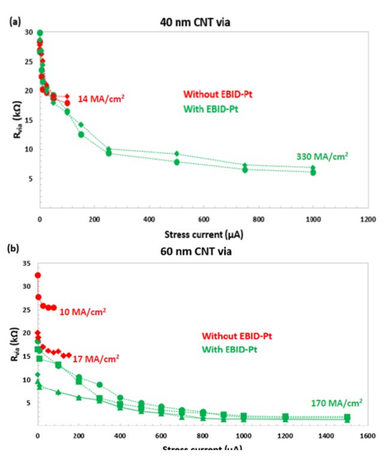


Figure 7. (a) Via resistance versus stress current behaviors for two 40 nm vias with and without Pt top contacts. Metallized top contact and current stressing form conformal interfaces between metal contact and the exposed CNT tips, resulting in lower via resistance, and improve the current-carrying capacity compared to unmetallized vias. (b) Similar results are shown for three 60 nm CNT vias with and without Pt top contacts

CNT/MLG

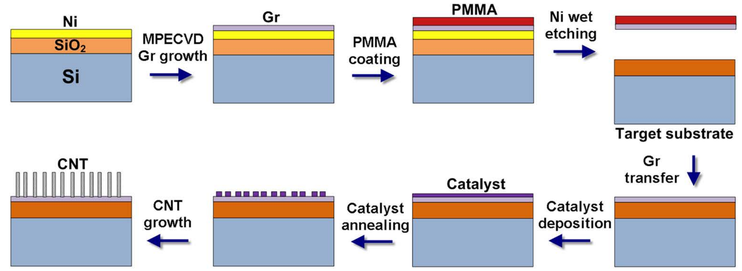


Figure 8. Process flow for CNT/MLG heterostructure fabrication [8]. MLG underlayer sputter-deposited with Ni catalyst film, followed by CNT growth in a plasma-enhanced chemical vapor deposition (PECVD) reactor, with various growth times and DC voltage.

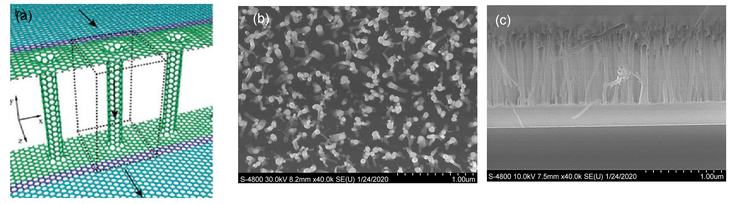


Figure 9. (a) Schematic of a 3D CNT/graphene interconnect structure. **(b)** Top-view SEM image of PECVD-grown CNT/MLG. **(c)** Cross-sectional SEM image of PECVD grown CNT/MLG. Imaging using scanning electron microscope (SEM) to obtain top and side-views images.

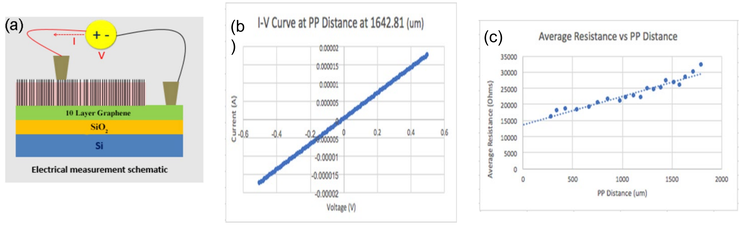


Figure 10. (a) Measurement schematic for CNT/MLG test structure [9]. **(b)** Typical I-V behavior for 500 VDC PECVD CNT/MLG. **(c)** Average resistance vs PP distance for 500 VDC PECVD CNT/MLG sample [10].

Current Findings: We have determined that certain CNT growth conditions (for example, reduced plasma voltage) improves the CNT/MLG interface by minimizing damage to the MLG underlayer, thus lowering the contact resistance and the resistance of the overall test structure.

References

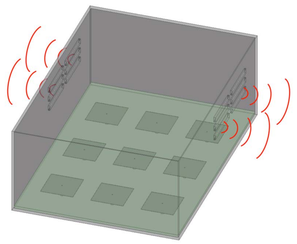
[1] Z. T˝okei, "End of Cu roadmap and beyond Cu," in Proc. 2016 IEEE Int. Interconnect Tech. Conf./Adv. Metallization Conf., 2016, pp. 1–58.
[2] International Technology Roadmap for Semiconductors, at http://www.itrs2.net/itrs-reports.html.
[3] M. A. Hussein and J. He, ""Materials' impact on interconnect process technology and reliability," *IEEE Trans. Semicon. Manufacturing* **18,** 69 - 85, 2005.
[4] A. Balandin, "Thermal properties of graphene and nanostructures carbon materials," *Nature Materials* **10**, 569-581, 2011.
[5] C. Zhou, A. A. Vyas, P. Wilhite, P. Wang, M. Chan, and C. Y. Yang, "Resistance Determination for Sub-100-nm Carbon Nanotube Vias," *IEEE Electron Device Letters* **36**, 71-73, 2015.
[6] A. A. Vyas, C. Zhou, and C. Y. Yang, "On-Chip Interconnect Conductor Materials for End-of-Roadmap Technology Nodes," *IEEE Trans. Nanotechnology* **17**, 4-10, 2018.
[7] A. A. Vyas, C. Zhou, Y. Chai, P. Wang, and C. Y. Yang, "Effect of improved contact on reliability of sub-60 nm carbon nanotube vias," *Nanotechnology* **27**, 375202 (11pp), 2016.
[8] C. Zhou, R. Senegor, Z. Baron Z., Y. Chen, S. Raju, A. Vyas, M. Chan, Y. Chai, and C. Y. Yang, "Synthesis and interface characterization of CNTs on graphene," *Nanotechnology* **28**, 054007 (10pp), 2017.
[9] R. Senegor and Z. Baron, "Carbon Nanotubes on Graphene: Electrical and Interfacial Properties," BSEE thesis, Santa Clara University, 2017.
[10] J. Shaffer and A. Michelmore, "Process Optimization for Carbon Nanotubes-on-Graphene Fabrication," BSEE thesis, Santa Clara University, 2018.

Design of Metamaterial Impedance Matching Surfaces at Near Field for EMC Solutions

Ali Khoshniat, PhD Candidate, Adviser: Dr. Ramesh Abhari
Department of Electrical and Computer Engineering, Santa Clara University
akhoshniat@scu.edu, rabhari@scu.edu

Santa Clara University
School of Engineering

Introduction

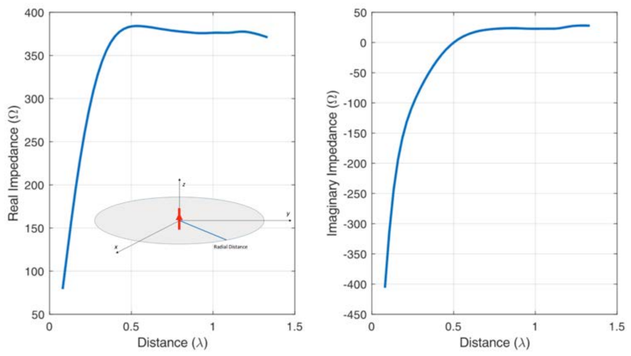


- Modern data centers, servers, routers, and storage units are running at tens of Gbps data rates. 25 Gig Ethernet and PCIe Gen5 (32 Gbps) are now commonplace examples of such systems.
- As a result, stronger radiated emissions are showing up in microwave and millimeter-wave frequency range.
- A custom-designed absorber is needed to target the frequency of electromagnetic compliance failure.
- The absorber acts as a **frequency selective medium** for maximum absorption of electromagnetic power. Its design principal follows **development of a metamaterial structure**.
- At the far field region the surface impedance of the proposed absorber should be 377 Ω at the failure frequency.
- In systems with electrically small chassis and metal enclosures, the distance between the source of radiation and the enclosure walls is in the near field region.
- A general methodology for designing this metamaterial absorber is developed in this work.

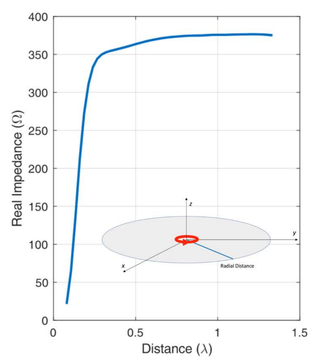
Wave Impedance Matrix

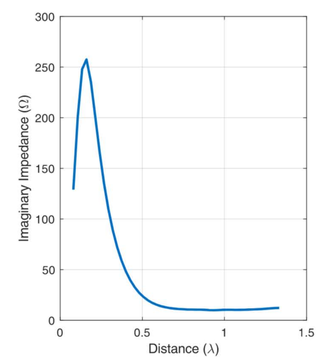
$$Z_{W\_spherical} \equiv \begin{bmatrix} \frac{E_r}{H_r} & \frac{E_r}{H_\theta} & -\frac{E_r}{H_\varphi} \\ -\frac{E_\theta}{H_r} & \frac{E_\theta}{H_\theta} & \frac{E_\theta}{H_\varphi} \\ \frac{E_\varphi}{H_r} & -\frac{E_\varphi}{H_\theta} & \frac{E_\varphi}{H_\varphi} \end{bmatrix} \qquad Z_{W\_Cartesian} \equiv \begin{bmatrix} \frac{E_x}{H_x} & \frac{E_x}{H_y} & -\frac{E_x}{H_z} \\ -\frac{E_y}{H_x} & \frac{E_y}{H_y} & \frac{E_y}{H_z} \\ \frac{E_z}{H_x} & -\frac{E_z}{H_y} & \frac{E_z}{H_z} \end{bmatrix}$$

Wave Impedance: Small Dipole Antenna

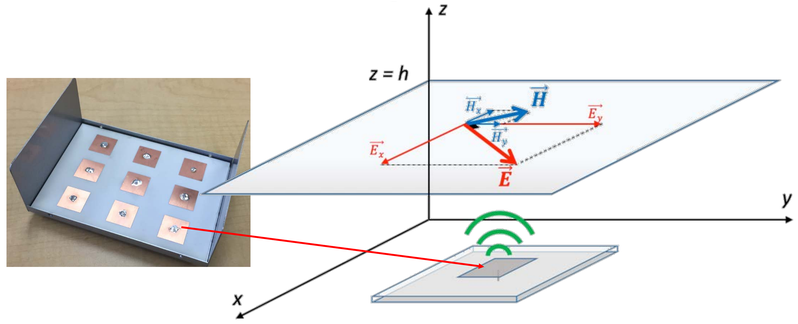


Wave Impedance: Small Loop Antenna



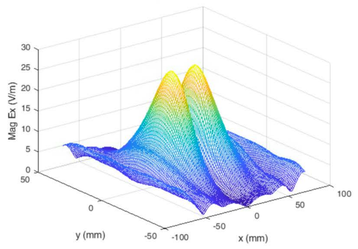


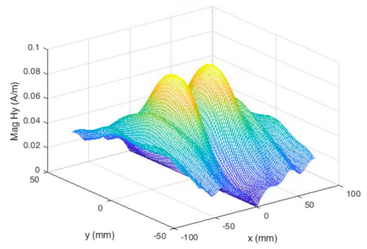
Base-Line Studied Structure: Square Patch



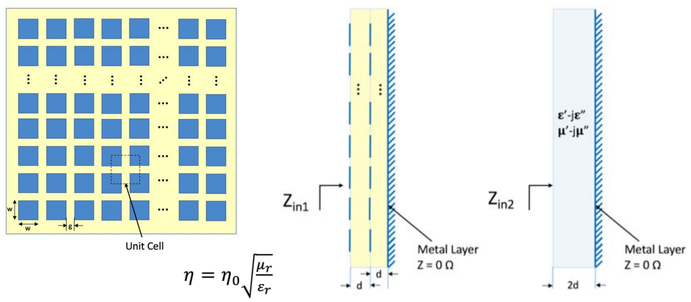
Hot spots of E and H fields for Impedance Calculation

- At 8 GHz, and h=10.94 mm distance (sample near field point), wave impedance at the hot spots for E and H fields is around 260-j100 Ω





Metamaterial Design: Equivalent Bulk Material Absorber



$$\eta = \eta_0 \sqrt{\frac{\mu_r}{\varepsilon_r}}$$

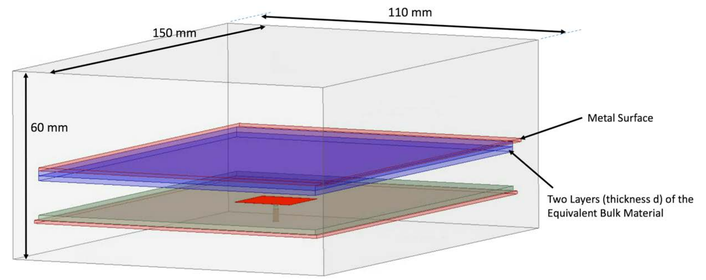
$$Z_{in2} = j\eta \tan\left(\frac{2\pi f}{c}\sqrt{\mu_r \varepsilon_r}\, l\right)$$

- Bulk material properties to achieve Z_{in2}= 377 Ω at 8 GHz (Far field)
  - ε_r = 5.1-j0.55
  - μ_r = 1.2-j0.95
- Bulk material properties to achieve Z_{in2}= 260-j100 Ω at 8 GHz (Near field)
  - ε_r = 3.7-j0.95
  - μ_r = 7.9-j0.55

Radiated Emissions Calculations

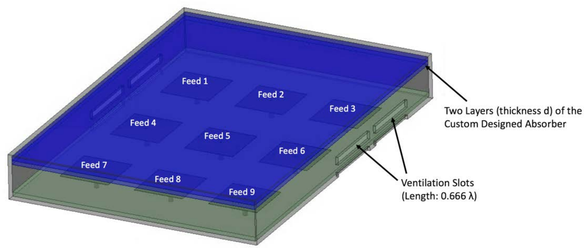
- A virtual box is defined around the radiation source
- Poynting Vector is integrated over each face
- Reflection from antenna port is considered
- Total radiated power calculated using:
  - Port reflection coefficient
  - Total integrated power from Poynting Vector

Studied Case	Radiated Power (mW)	Reflection Coefficient (dB)	Effective Radiated Power (mW)
1- Antenna Alone (Baseline)	7.715	-7.13	13.776
2- Antenna Covered with a Copper Plate	8.710	-8.89	13.524
3- Antenna Covered with 377 Ω Absorber	0.380	-6.65	0.710
4- Antenna Covered with 260+j100 Ω Absorber	0.923	-5.64	1.728
5- Antenna Covered with 260+j100 Ω Absorber	0.251	-5.80	0.512
6- Antenna Covered with Microwave Absorber	2.827	-6.92	5.149



A Mock-up High-System with Metal Enclosure Box

- 3x3 Array to simulate worst case scenarios of radiation sources
- Distance of the top wall of the chassis from system board h = 10.94 mm (0.3λ)
- Emission through ventilation slots
- Comparison with commercial microwave absorber
- Relative Radiated Power calculated for each case $RRP = 10 \times Log_{10}\left(1 - \left(\sum_{j=1}^{9} |S_{2j}|^2\right)\right)$



Active Feed Location	No Absorber			377 Ohm			260-j100 Ohm			260+j100 Ohm			Microwave Absorber		
	Max rE (dBV/m)	RRP (dB)	Max Effective rE (dBV/m)	Max rE (dBV/m)	RRP (dB)	Max Effective rE (dBV/m)	Max rE (dBV/m)	RRP (dB)	Max Effective rE (dBV/m)	Max rE (dBV/m)	RRP (dB)	Max Effective rE (dBV/m)	Max rE (dBV/m)	RRP (dB)	Max Effective rE (dBV/m)
Feed 1/3	-4.5	-6.3	1.8	-8	-1.1	-6.9	-5.7	-1	-4.7	-9.7	-1.4	-8.3	-9.8	-2.5	-7.3
Feed 2	0.7	-4.3	5	-18.9	-1	-17.9	-12.5	-1	-11.5	-19.1	-1.2	-17.9	-8.5	-1.7	-6.8
Feed 4/6	-3.3	-9.6	6.3	-8.9	-2.1	-6.8	-7.5	-2.1	-5.4	-10.3	-3	-7.3	-9.4	-6.8	-2.6
Feed 5	-0.8	-5.2	4.4	-19.9	-1.1	-18.8	-13.5	-1.2	-12.3	-20.2	-1.4	-18.8	-10.4	-2.2	-8.2
Feed 7/9	-1.6	-5.8	4.2	-29.8	-1.2	-28.6	-25.7	-1.3	-24.4	-30.9	-1.5	-29.4	-20.6	-2.7	-17.9
Feed 8	-2.5	-4.6	2.1	-26.6	-1	-25.6	-20.7	-1	-19.7	-25.9	-1.2	-24.7	-13.6	-1.8	-11.8

Conclusions

- A general methodology based on wave impedance calculations is proposed to design absorbers for high-speed/high-frequency systems with metal enclosures.
- The proposed absorber is designed based on metamaterial theory to exhibit an input surface impedance that yields maximum absorption of emitted electromagnetic power.
- Study of a number of test cases proves the effectiveness of the proposed solution demonstrating its superior performance in comparison to that of a commercial microwave absorber.

Efficient Hardware/Software Co-design Partitioning of WLANs for Software-defined Radio

Rami Akeela and Yacoub Elziq
Department of Electrical and Computer Engineering, Jimzu Technologies

Introduction

Co-design fundamentally is the process of partitioning the system into synthesizable hardware and executable software blocks. System developers have the responsibility of setting strict rules or a criteria in order to determine the partitioning process.
Finding the pseudo-optimal design partition is not trivial. In fact, most HW/SW partitioning algorithms have proven to be NP-hard (non-deterministic polynomial-time hard).

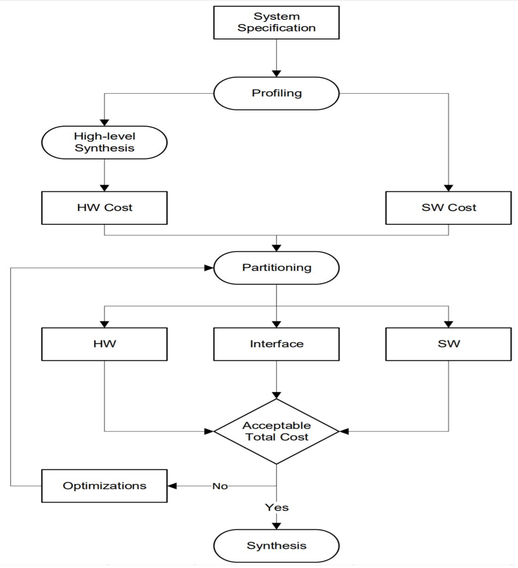
Objectives

An SDR domain-specific partitioning methodology in conjunction with a set of design guidelines, which aim to generate a performance enhancing HW/SW partitioning of WLANs.
This is done through the use of a formal systematic approach and a cost function that helps calculate the health factor of partitions by examining several design parameters and constraints including power, timing, and resource utilization.

Definitions

FPGA = Field Programmable Gate Array. Can be thought of as a dynamic ASIC.
ASIC = Application Specific Integrated Circuit.
HW = Hardware. Generally meaning what runs on the FPGA fabric.
SW = Software. Generally meaning what runs on the CPU.
HLS = High-level Synthesis.
WLAN = Wireless Local Area Network.
PHY = Physical Layer in WLAN protocols.
FFT = Fast Fourier Transform.
AXI = the Advanced eXtensible Interface.
DMA = Direct Memory Access.
FIFO = First In First Out.
LUT = Look-up Table.
FF = Flip Flop.

Partitioning Methodology



Factors for efficient partitioning:

- Target architecture
- Complexity of the system to be partitioned
- Partitioning decision
- Granularity of system components
- Estimation method
- Metrics used to calculate the partition cost
- Optimization goal: performance, power, and area
- Optimization method

Results

We implement the IEEE 802.11a PHY receiver chain using our partitioning methodology on the Zynq ZC706 SoC using an HLS tool, namely Xilinx SDSoC, which enables us to generate software executables, hardware RTL, and data movers. We achieve an average speedup of 8.7x in comparison to the all-software implementation. Also, with the set of optimization techniques adopted, we are able to further accelerate the co-design by a factor of 1.15x.

Design Scheme	Hardware Cost C_total^HW (10^{-3})	Software Cost C_total^SW (10^{-3})	Power Consumption P_total^HW (mW)	Execution Time T_total (ns)	Memory Cost M_total (KB)
All-SW	–	346	–	342	239
Co-design 1	974.2	48.1	484	45	154
Co-design 2	986.1	48.1	497	39	175
All-HW	1174.2	–	531	42	–

Component	Hardware Cost C_total^HW (10^{-3})	Software Cost C_total^SW (10^{-3})	Power Consumption P_total^HW (mW)	Execution Time T_total (ns)	Memory Cost M_total (KB)
Scrambler	7.53	2.76	6	2.8	17.2
Encoder	8.24	2.77	9	3	19.1
Decoder	297.3	15.4	33	9.7	58.8
Interleaver	265	16.3	27	8.9	46
OFDM(FFT)	282.5	13.2	31	10.1	49.7

Pseudo-optimization

When targeting the Xilinx Zynq-7000 SoC, developers have the choice to implement functions either on processors or on FPGA using multiple hardware data movers like DMA and FIFO buffers with three different types of interfaces for moving data between the FPGA and processor, such as AXI-Lite and AXI-Stream. This results in at least 16 possible configurations to optimize per function, which could be very time-consuming to evaluate using the traditional embedded design methodologies. Given the available hardware resources (Xilinx ZedBoard SoC), a constrained pseudo-optimized solution is generated. This is done through the use of Gradient Descent, which is an iterative optimization algorithm, used to find the minimum value for a function. The general idea is to initialize the parameters to random values, and then take small steps in the direction of the "slope" at each iteration.

$$Cost = 100 \times \left(\frac{T_{total}}{T_{max}} \times \beta_T + \frac{C^{SW}_{total}}{C^{SW}_{max}} \times \beta_{SW} + \frac{C^{HW}_{total}}{C^{HW}_{max}} \times \beta_{HW} + \frac{P_{total}}{P_{max}} \times \beta_P + \frac{M_{total}}{M_{max}} \times \beta_M \right)$$

Further Discussion

Communication to FPGA is split into two main methods:

A Based on commands that are formatted with a header, followed by optional data (inputs for the command). FPGA sends replies to SW after a command is completed or in the case of any errors. These are sent over an AXI4-stream interface.

B Using an instruction memory and data register approach, SW has direct access to FPGA memory and can configure more complex logic flows. These are sent over an AXI4-lite interface.

Depending on the FPGA board used, communication is exposed to SW through a C++ library over USB-UART. There are wrappers that convert the communication method to the internal FPGA AXI-lite and AXI-stream interfaces.
Data Motion Network Resource Utilization:

Resource	Used	Total	% Utilization
Clocking	4	32	12.5
I/O	34	128	26.56
LUT	8096	218600	3.7
FF	4283	437200	0.98

References

[1] Rami Akeela and Yacoub Elziq. Design and Verification of IEEE 802.11ah for IoT and M2M Applications. *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 491–496, Mar 2017.

[2] Rami Akeela and Behnam Dezfouli. Software-defined Radio: Architecture, State-of-the-art, and Challenges. *Elsevier Computer Communications*, pages 106–125, 2018.

[3] Rami Akeela and Yacoub Elziq. Efficient co-design partitioning of wlans on soc-based sdrs. *Microsystem Technologies*, pages 1–18, 2019.

Contact Information

Author 1 E-mail rakeela@scu.edu

Author 2 E-mail elziq.ip@gmail.com

Optimal Energy Management of a University Campus Microgrid Integrating Solar Energy and Fuel Cell

Sowmya Bellam, Isaac Marcia, and Johann Espinosa
Advisor: Dr. Maryam Khanbaghi
Department of Electrical and Computer Engineering
Santa Clara University
School of Engineering

BACKGROUND

Microgrid

The new generation microgrids not only supply backup power, but also a more complex configuration that contains all the essential elements of a large-scale grid such as the ability to balance electrical demand with sources, schedule the dispatch of resources, and preserve grid reliability. Therefore, it is easier to convert large, consistent loads such as university/office campuses into a new generation microgrid.

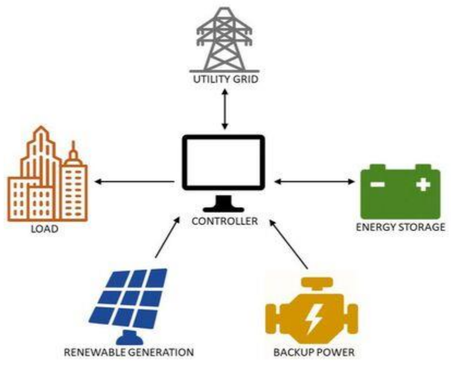


Figure 1 : Schematic of an energy management system of a microgrid

Introduction

The usage of Renewable Energy (RE) has become a global priority for two major reasons: environmental impact of fossil fuel and its future scarcity. Intense research is being conducted on the integration of Renewable Energy Sources (RES) into existing power generation systems. One proposed solution revolves around the concept of microgrid. Microgrids are becoming increasingly popular, combining multiple energy generation and storage systems to diversify and secure their private energy portfolios.

Objective

The goal of this work is to explore optimal energy management of a university campus integrating solar energy, storage and fuel cell. Energy management system (EMS) of the university campus considered here must meet three requirements:

- Satisfy campus load
- Use sustainable energy
- Minimize electricity costs

To achieve the objective, this work is divided into three parts-

- System Description
- Method
- Observations and Results

SYSTEM MODEL

System Description

The campus electric power is provided by different energy sources. They are comprised of different types of roof-top photovoltaic panels (different size and efficiency), solar thermal systems and a fuel cell, Fuel cell installed also has some storage. These are connected to the local distribution system via two different transformers where each serves half of the campus. A Tesla energy storage system is also under consideration.

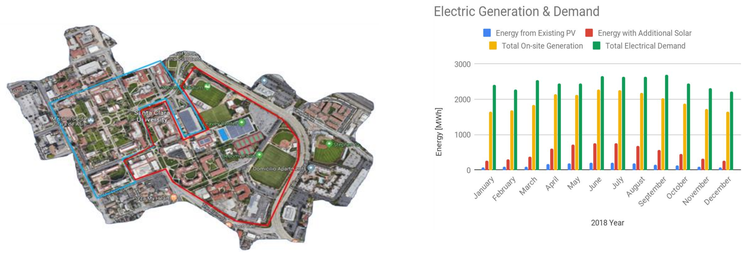


Figure 2 : SCU electrical generation coverage

Method

For this work we envisioned the campus as two connected microgrids. Our goal for this two connected microgrids system is that, at any point in time, each can be disconnected from the grid or the other microgrid. Managing both microgrids simultaneously by overcoming the intermittent presence of solar energy and meeting the above-mentioned requirement is a challenging problem.

There have been many studies on energy management of smart buildings integrated with renewable energy and EMS of microgrids. Considering commercial spaces such as university/office campuses are steady loads compared to residential loads, managing energy provided to these loads is easier compared to a random residential load.

We have performed load analysis and differentiated load according to the type of the building usage. Figure 3, compares SCU load data with an aggregated commercial data provided by NREL.

We combined lecture halls and dining halls power consumption since they have a comparable load pattern. Similarly, student dorms and libraries loads are combined together. Since the university has several parking structures, we compared rooftop solar design with carport design. In terms of design, construction cost and efficiency, it was observed that carports would suit the requirements. In the system we therefore used carport design for most of the parking structures.

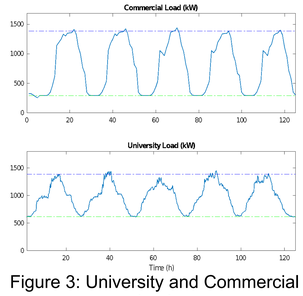


Figure 3: University and Commercial load

RESULTS & RESEARCH

Method Cont.

Using this new design improves usage of parking spots. The examined parking lots have very few shading losses, hence carports can efficiently draw energy. Also, location of examined spots are near to the lecture halls; therefore, power can be distributed among the nearby loads. For roofs with good direction and less HVAC disturbances we used rooftop solar as well.

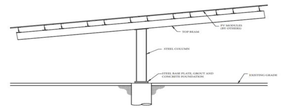


Figure 4: Solar Structure

Observations and Results

Performing analysis using a simulation tool, PVsyst allowed us to experiment with the above-mentioned structure on various locations. It was observed through the simulation that the net impact of solar and fuel cells could satisfy up to 74% of SCU's electricity demand over a year.

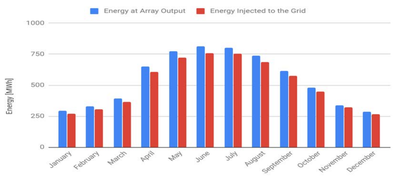


Figure 5: Energy flow

Conclusions & Future Works

The focus of this project was on the usage of microgrids, their advantages, and promoting their usage in commercial spaces. The proposed analysis was presented to the SCU energy operation team. Additionally, using these spaces can educate young professionals on the importance of energy efficiency in today's world.

Our future work will focus on EMS of the SCU campus, two microgrids when assumed as networked microgrids in islanded mode, and compare the two configurations from efficiency and cost perspectives. Furthermore, we will analyze different types of fuel cells according to the on-site generation needs and storage. Fuel Cell will be considered as a backup generation source. Another important step in the future of this project would be sizing and integration of battery storage into the buildings on campus. This would be achieved by analyzing smart meter data from buildings on campus and analyzing their peak energy usage and consumption patterns.

Acknowledgment

Authors would like to thank the School of Engineering, Silicon Valley Power, the SCU Operations team, and the Latimer Energy lab for their support.

Reconfigurable Yagi-Uda Antenna Arrays with Passive Beamforming for Microwave Hyperthermia Applications

Nivedita Parthasarathy, PhD Student; Dr. Ramesh Abhari, Advisor
Department of Electrical and Computer Engineering
Santa Clara University
School of Engineering

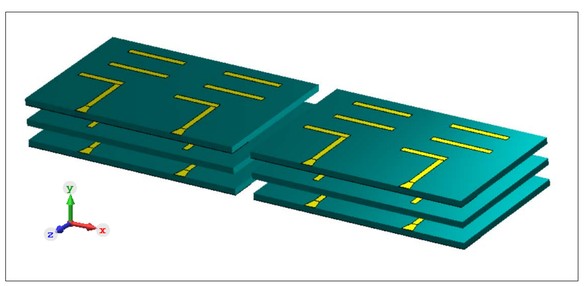
Motivation:

- Antenna arrays are often used as applicators in thermo-therapeutic applications to focus electromagnetic energy in a specific region of the human body.
- The use of a 2D conformal array with beamforming network allows the antenna setup to switch or steer the focused beam in the desired region of treatment.
- We present two configurations of Yagi antenna arrays for possible use in microwave hyperthermia applications. A 1x4 collinear array is designed and fed with a metamaterial Butler Matrix feed network to provide beam switching at four specific directions.

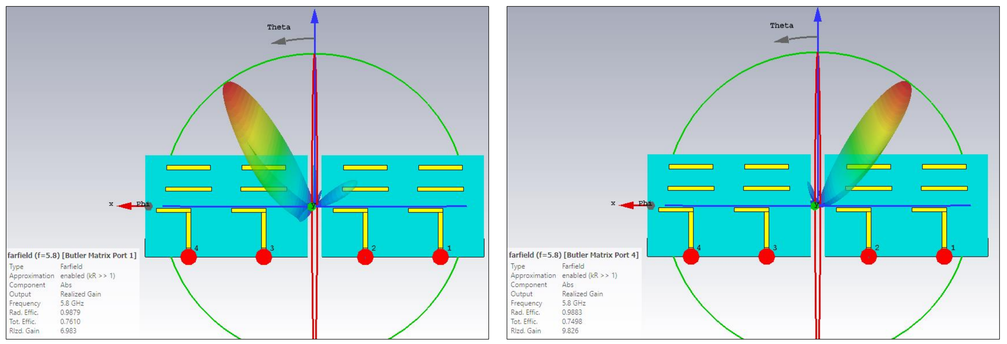
Objectives:

- A 2x2 Yagi antenna array previously developed is investigated in two scenarios at 5.8 GHz frequency of operation.
- To evaluate the 2D array's efficacy as a hyperthermia applicator, an electromagnetic model of the human body in CST Studio Suite, "Nelly" containing skin, fat and muscle-fat averaged tissues is used for simulation.

3D View of Stacked Yagi Array:



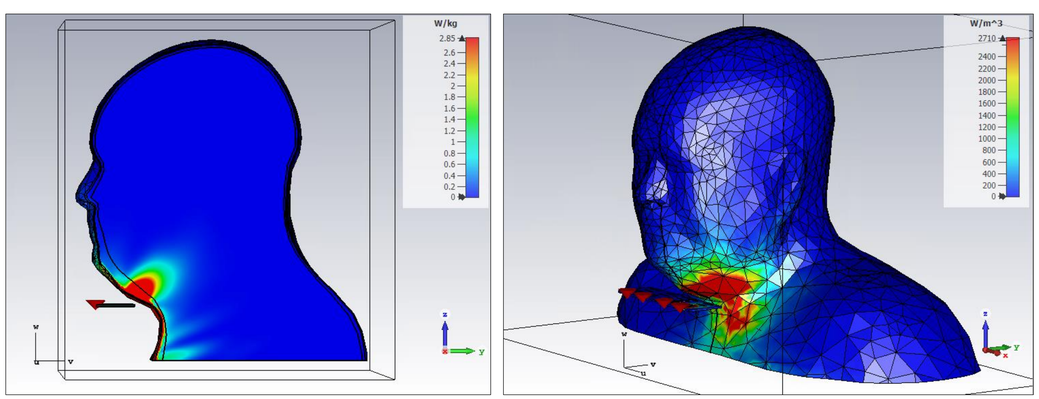
Simulation Results Demonstrating Beamforming with Butler Matrix



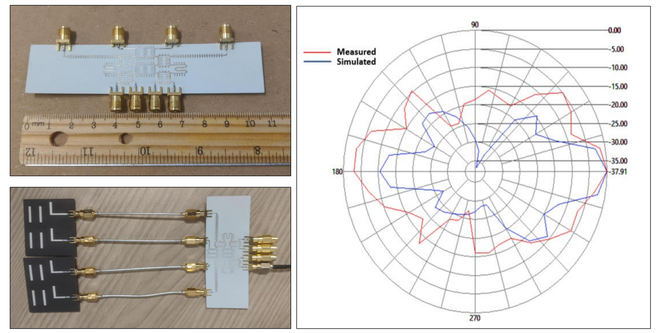
Comparison of Array Performance at 5.8 GHz

Yagi Antenna Arrangement	Antenna Performance			
	Isolation	*Peak Gain*	*HPBW ($\phi = 0°$)*	*HPBW ($\phi = 90°$)*
1x2 Stacked Planar Array (d_y=0.5λ)	17.328 dB	9.855 dBi	62.6°	77.7°
2x2 Stacked Yagi Array (d_x=0.5λ and d_y=0.2λ)	5.48 dB	6.45 dBi	43°	53.7°
2x2 Stacked Yagi Array with Parasitic Coupling (d =0.5λ and d_y=0.2λ)	18.8 dB	8.787 dBi	41.7°	50°
1x4 Collinear Yagi Array (d_x =0.5λ)	14.92 dB	11.7 dBi	20.9°	96.5°
2x4 Stacked Yagi Array with Parasitic Coupling (d_x =0.5λ and d_y =0.2λ)	15.05 dB	13.09 dBi	20.7°	71.8°

Full-wave Analysis with Biomodel in CST Studio Suite



Measurement of Collinear Array:



Result Findings and Future Work:

- The collinear 1D array was studied for its capability to provide beam switching when fed by a Butler Matrix feed network. The design was confirmed via measurements.
- Simulation results including human body model suggest the feasibility of this design as a hyperthermia applicator.
- Current studies underway to investigate hyperthermia treatment planning by means of a HyperCollar – a conformal arrangement of the stacked antennas to maximize radiation to target area.

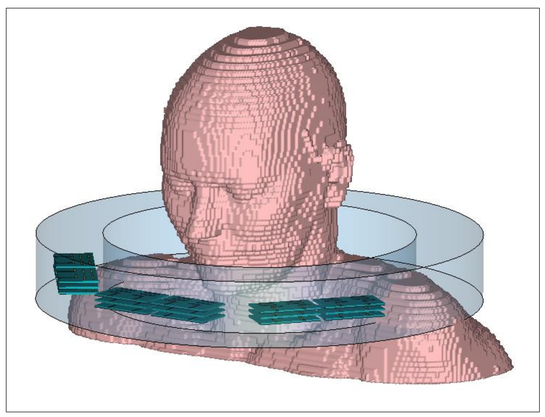


Illustration of proposed HyperCollar setup

Autonomous Multi-Robot Exploration of Planetary Caves

Kamak Ebadi - Ph.D. Candidate
Thesis Advisor: Dr. Sally Wood
Department of Electrical and Computer Engineering

Jet Propulsion Laboratory
California Institute of Technology

Santa Clara University
School of Engineering

BACKGROUND AND MOTIVATING APPLICATION

Planetary subsurface voids (lava tubes, cave) are of enormous importance:

- ✓ Orbiters, rovers and landers only map the surface of planetary bodies
- ✓ Planetary caves provide:
  - ➢ valuable information about planet's history and evolution
  - ➢ ideal conditions for the development of microbial life
  - ➢ potential habitats for future human space missions

CHALLENGES

- ➢ GPS : unavailable
- ➢ Inertial sensors: noisy and unreliable
- ➢ Wheel : uneven and slippery terrains
- ➢ Vision : darkness, dust, non-Lambertian surfaces
- ➢ Lidar : self-similar, flat or featureless surfaces

PROBLEM FORMULATION

In graph-based SLAM we model the map obtained by a robot as a graph:

$$\mathcal{G}^\alpha \langle \mathcal{X}^\alpha, \mathcal{U}^\alpha, \mathcal{Z}^\alpha, \mathcal{O}^\alpha \rangle$$

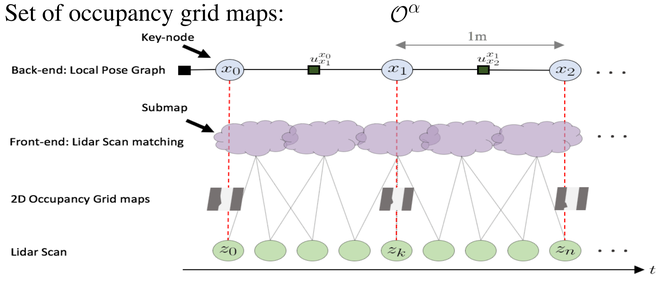
Trajectory representing a finite set of poses

$$\mathcal{X}^\alpha = \{x_1^\alpha, x_2^\alpha, \cdots\}\ ,\ x_i^\alpha = (R_i^\alpha, t_i^\alpha)$$

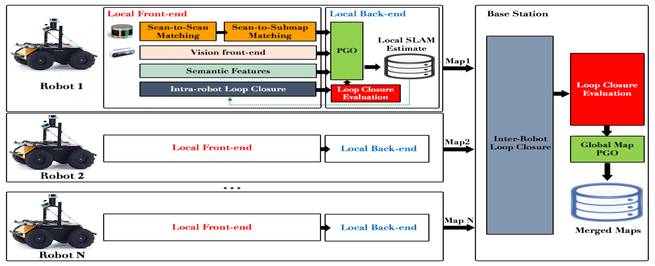
Set of relative motions between poses: $\mathcal{U}^\alpha$

Set of lidar measurements: $\mathcal{Z}^\alpha$

Set of occupancy grid maps: $\mathcal{O}^\alpha$



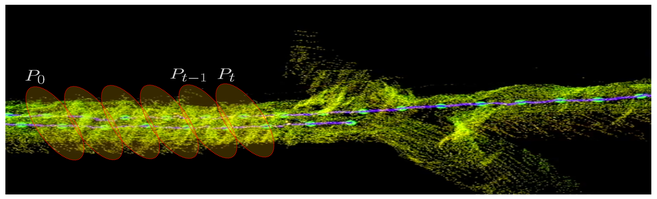
PROPOSED METHOD



1- Point Cloud Filtering

- ➢ 3D Feature detection and extraction

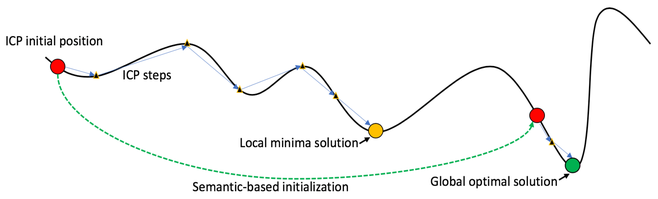
2- Lidar Scan Matching - Odometry



$$u_t^{t-1} = \underset{\hat{\mathcal{T}}}{\operatorname{argmin}}\, \mathcal{E}(P_{t-1}, \hat{\mathcal{T}}(P_t^*)) = \frac{1}{N}\sum_{i=1}^{N} \left\| P_{t-1} - \hat{\mathcal{T}}(P_t^*) \right\|^2$$

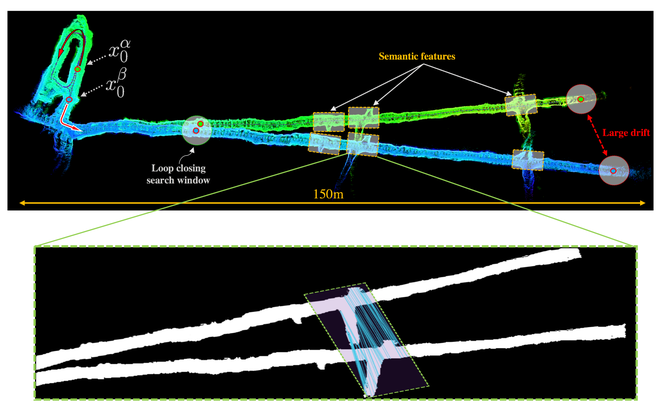
3- Semantic-Geometric Loop Closure Detection

- ➢ Two-stage optimization process:



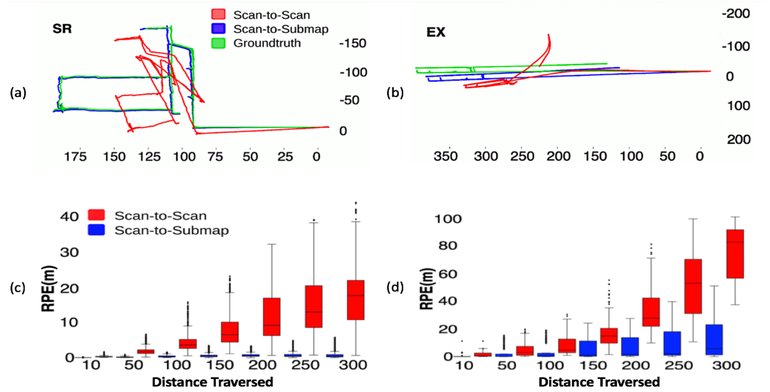
$$\mathcal{T}^0 = \underset{T}{\operatorname{argmin}}\, \mathcal{D}(F_i, TF_j)$$

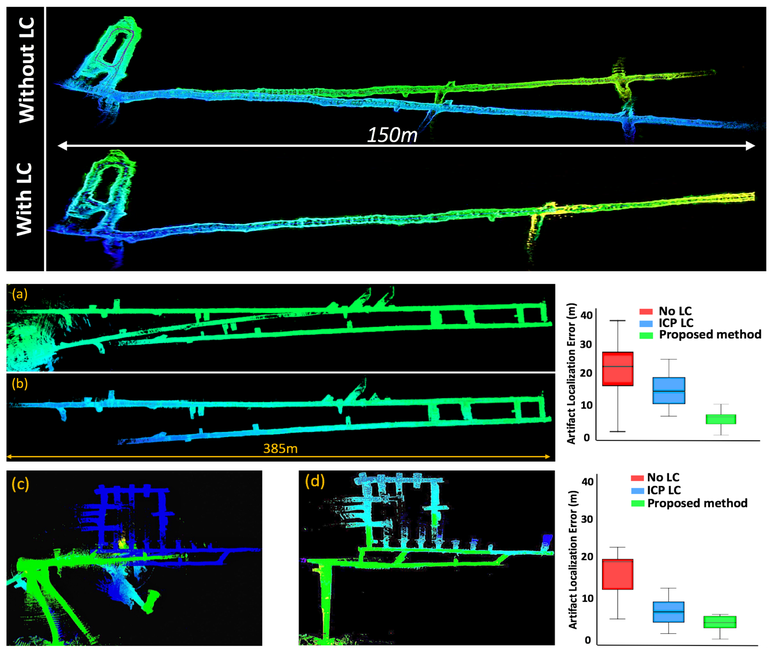
$$\mathcal{T}^* = \underset{\mathcal{T}}{\operatorname{argmin}}\, \mathcal{E}(P_i, \mathcal{T}P_j, \mathcal{T}^0)$$



EXPERIMENTAL RESULTS

Two autonomous robots deployed 1000ft underground at the Arch mine





ACKNOWLEDGEMENT

This research has been carried out at NASA Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

Department of Electrical and Computer Engineering

A Deep Learning-enabled Guide for the Visually Impaired

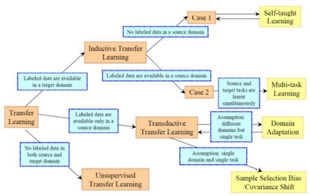
Allen Shelton[1], Prof. Tokunbo Ogunfunmi[1]
1 Santa Clara University, 500 El Camino Real, Santa Clara, 95053

ABSTRACT AND MOTIVATION

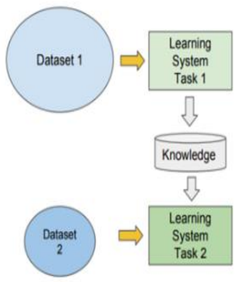
Machine learning has become a very popular field of research over the past several years as artificial intelligence seeks to find more ways to automate different tasks using data. Deep learning, a subset of machine learning, uses artificial neural networks to classify data, and is a very popular method for image recognition. I think that deep learning can be used to improve the way of life of visually impaired people. There are 1.3 billion people globally with some sort of vision impairment, and 36 million of that group are legally blind. Normally, in the absence of assistance from other people, blind people will use guide dogs when travelling outside their homes. Even though guide dogs have been and currently are of great use to blind people, they not only can be expensive, but they don't give them very much information on where they are or what's around them. If there were a way for people with severe visual impairments to not only avoid hazards, but also know what's around and how to navigate their environment, not only would their safety increase, but their way of life would be significantly improved. This project seeks to use MATLAB to develop an application to carry out this task. This application would use a camera that a user would point at different objects, buildings, or landmarks, and identify what it sees. Then the application would utilize a speaker to verbalize what is identified so that blind people will know what is around them. Convolutional neural networks (CNN) are very effective machine learning algorithms for image identification, and MATLAB gives users the abilities to create them relatively more simply than other languages, either through code or through its Deep Network Designer application. I specifically want to use transfer learning for my project, where I take a preexisting neural network, namely AlexNet, and tweak it to my unique specifications in order to solve this specific problem. I would also employ a text-to-speech function in order to verbalize what the network identifies.

TRANSFER LEARNING

- Getting large amounts of labeled data to train CNNs can be very difficult
- Traditional Machine Learning problem solving is isolated and doesn't take from similar previously solved problems
- Use knowledge from pretrained models to solve new problems
- Faster learning process, higher accuracy, and less training data required

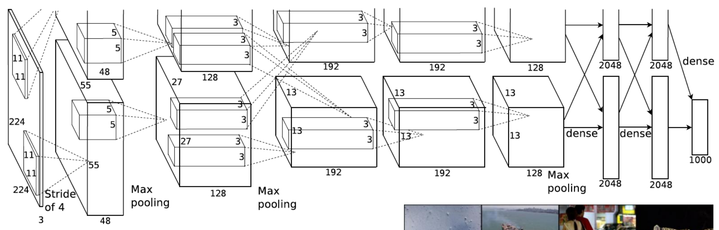


Transfer Learning

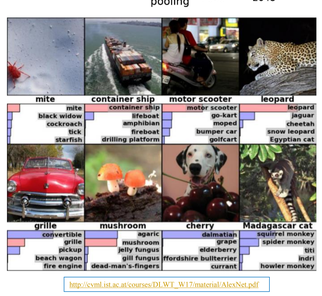


https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a

ALEXNET/IMAGENET



- 5 convolutional layers and 3 fully connected layers.
- Also has max pooling and dropout layers
- Requires image input size of 227 x 227
- 63.2 million parameters and 650,000 neurons
- Took 5 to 6 days to train
- Winner of ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012
- Trained on ILSVRC dataset - 1000 categories
  - 1 million training images, 50000 validation images, 150000 test images



http://cvml.ist.ac.at/courses/DLWT_W17/material/AlexNet.pdf

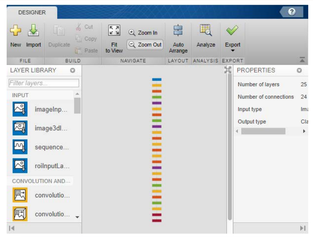
• A. Krizhevsky, I. Satskever, G.E. Hinton, *ImageNet Classification with Deep Convolutional Neural Networks*, 2012

THE DATASET

The dataset for this project is a mixture of images found from the Internet and images taken from my smartphone. The initial vision for this application will be to assist a blind student in navigating Santa Clara University, so the categories of objects that the CNN will identify have been broken up into 2 groups, each to be trained separately. One group is focused on objects that any student may commonly encounter on campus or within any of SCU's many buildings. The other categories are a select number of the buildings themselves. A student would not only be aware of their surroundings within their dorm, but also know what buildings they are near when they are outside.

Group 1	Group 2
Bed	Benson Center
Bench	Bronco Statue
Car	Casa Italiana RLC
Couch	Daly Science
Door	Dowd Art and Art
Elevator	History Building
Exit Sign	Graham RLC
Fire Extinguisher	Kenna Hall
Folder/Binder	University Library
Fork	Mission Church
Knife	O'Connor Hall
Palm Tree	
Pen/Pencil	
Stairs	

MATLAB'S DEEP NETWORK DESIGNER



AlexNet Architecture visualized in Deep Network Designer

With MATLAB's Deep Network Designer application, one can easily create and edit neural networks through a graphical user interface instead of with code. It includes functionality to add and edit the properties of network layers. Additionally, the application can generate MATLAB code from the networks that are built within it. Since the pretrained structure and parameters are available through the Deep Learning Toolbox, AlexNet can be imported into Deep Network Designer and modified so that transfer learning can be used on my dataset.

RETRAINING

After loading all of the images into named subfolders, training can begin. First, I call the imageDatastore function in order to load the images into MATLAB for training, and I set the labels for the images to be the names of the subfolders they were in. I split up the images so that 70% were for training and 30% is for validation. I used the imageDataAugmenter function to randomly rotate and translate the images during training to avoid overfitting of the network to the training data. Additionally, MATLAB offers several training options users can modify that affect the training process, including initial learn rate, maximum number of epochs, and mini-batch size. Those options are shown here. Finally, the trainNetwork function takes the image data, the layers of AlexNet, and the training options, and begins retraining the network

```
inputSize = alex.Layers(1).InputSize;
allImages = imageDatastore('myImages', ...
    'IncludeSubfolders', true, ...
    'LabelSource', 'foldernames');
[trainingImages, valImages] = splitEachLabel(allImages, 0.7, 'randomized');
imageAugmenter = imageDataAugmenter( ...
    'RandRotation',[-20,20], ...
    'RandXTranslation',[-3 3], ...
    'RandYTranslation',[-3 3]);
augimgTrain = augmentedImageDatastore(inputSize(1:2),trainingImages, ...
    'DataAugmentation',imageAugmenter);
augimgValidation = augmentedImageDatastore(inputSize(1:2),valImages);
```

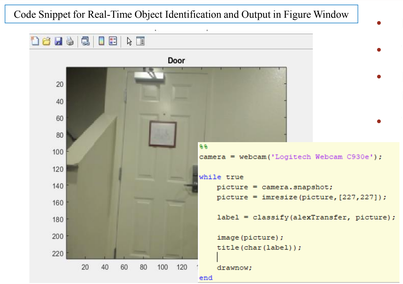
Loading and Preprocessing of Image Data

```
options = trainingOptions('sgdm', ...
    'InitialLearnRate', 0.0001, ...
    'MaxEpochs', 10, ...
    'MiniBatchSize', 30, ...
    'ValidationData',augimgValidation, ...
    'ValidationFrequency',2, ...
    'ValidationPatience',10, ...
    'LearnRateSchedule','piecewise', ...
    'LearnRateDropPeriod',4, ...
    'Plots','training-progress', ...
    'Verbose',true, ...
    'VerboseFrequency',20);
alexTransfer = trainNetwork(augimgTrain, layers, options);
```

Training Specification of CNN

REAL-TIME OBJECT IDENTIFICATION

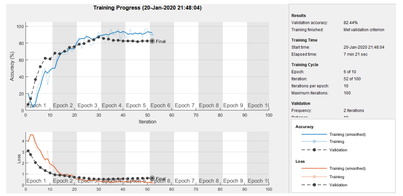
Code Snippet for Real-Time Object Identification and Output in Figure Window



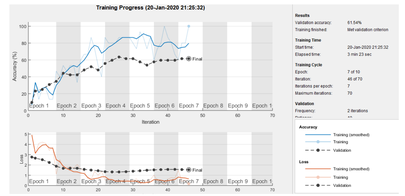
- MATLAB offers a support package for USB webcams
- This allows you to bring live images into MATLAB
- I can combine this functionality with the retrained CNN to identify objects seen in the live feed of the webcams
- To do this, I simply created a forever loop that:
  - Captures an image from the webcam
  - Classifies the image with the neural network
  - Displays the image in a figure window with the title being the classification from the network.

RESULTS OF NETWORK TRAINING

Group 1



Group 2

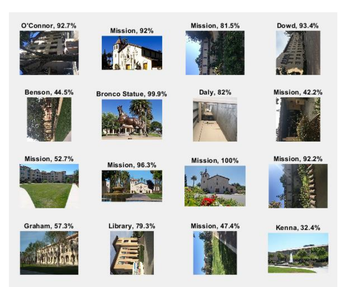


Validation Data Accuracy/Loss

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Validation Accuracy	Mini-batch Loss	Validation Loss	Base Learning Rate
1	1	00:00:14	20.00%	6.67%	3.9187	3.1084	1.0000e-04
1	2	00:00:27	10.00%	13.74%	4.5743	2.7623	1.0000e-04
1	4	00:00:43	13.33%	36.64%	3.4554	2.0309	1.0000e-04
1	6	00:00:59	33.33%	51.91%	3.1172	1.5911	1.0000e-04
1	8	00:01:16	50.00%	61.83%	2.2002	1.3239	1.0000e-04
1	10	00:01:31	66.67%	61.07%	1.4356	1.1947	1.0000e-04
2	12	00:01:49	50.00%	67.94%	1.4818	0.9255	1.0000e-04
2	14	00:02:06	70.00%	67.18%	1.0523	0.9663	1.0000e-04
2	16	00:02:22	66.67%	70.23%	0.9740	0.8946	1.0000e-04
2	18	00:02:38	73.33%	74.81%	0.8833	0.7690	1.0000e-04
2	20	00:02:53	73.33%	75.57%	0.7241	0.7072	1.0000e-04
3	22	00:03:12	86.67%	70.99%	0.8310	0.8722	1.0000e-04
3	24	00:03:28	83.33%	79.39%	0.3693	0.6365	1.0000e-04
3	26	00:03:44	80.00%	82.44%	0.7853	0.5551	1.0000e-04
3	28	00:04:01	80.00%	83.97%	0.5300	0.5506	1.0000e-04
3	30	00:04:16	93.33%	87.02%	0.2025	0.5339	1.0000e-04
4	32	00:04:34	90.00%	85.50%	0.2742	0.5289	1.0000e-04
4	34	00:04:51	86.67%	86.73%	0.3224	0.5369	1.0000e-04
4	36	00:05:08	90.00%	83.21%	0.4194	0.5528	1.0000e-04
4	38	00:05:24	96.67%	83.21%	0.2030	0.5613	1.0000e-04
4	40	00:05:39	93.33%	82.44%	0.1488	0.5663	1.0000e-04
5	42	00:05:58	86.67%	82.44%	0.3417	0.5770	1.0000e-05
5	44	00:06:14	90.00%	82.44%	0.4401	0.5919	1.0000e-05
5	46	00:06:31	90.00%	81.68%	0.3195	0.6059	1.0000e-05
5	48	00:06:47	96.67%	82.44%	0.3965	0.6142	1.0000e-05
5	50	00:07:03	96.67%	82.44%	0.1914	0.6231	1.0000e-05
6	52	00:07:21	90.00%	82.44%	0.2268	0.6234	1.0000e-05

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Validation Accuracy	Mini-batch Loss	Validation Loss	Base Learning Rate
1	1	00:00:23	6.67%	9.92%	6.9192	2.7675	1.0000e-04
1	2	00:00:33	33.33%	23.05%	3.1259	2.6611	1.0000e-04
1	4	00:00:43	26.67%	25.00%	4.2164	2.5125	1.0000e-04
1	6	00:00:51	13.33%	30.77%	3.4136	2.2397	1.0000e-04
2	8	00:01:59	46.67%	34.62%	1.1143	1.8775	1.0000e-04
2	10	00:02:06	53.33%	44.23%	1.4179	1.6690	1.0000e-04
2	12	00:02:16	53.33%	52.31%	1.8528	1.4802	1.0000e-04
2	14	00:02:25	33.33%	52.31%	2.0514	1.4817	1.0000e-04
3	16	00:02:29	66.67%	48.00%	1.4379	1.5764	1.0000e-04
3	18	00:02:34	66.67%	51.92%	0.9391	1.5016	1.0000e-04
3	20	00:02:46	53.33%	48.00%	1.0013	1.4399	1.0000e-04
4	22	00:02:55	66.67%	55.77%	0.5309	1.3640	1.0000e-04
4	24	00:02:00	80.00%	59.62%	0.6023	1.3224	1.0000e-04
4	26	00:02:07	86.67%	63.46%	0.4742	1.3222	1.0000e-04
4	28	00:02:16	86.67%	61.54%	0.5835	1.3471	1.0000e-04
5	30	00:02:23	93.33%	59.62%	0.3906	1.3968	1.0000e-05
5	32	00:02:30	93.33%	57.69%	0.1401	1.4301	1.0000e-05
5	34	00:02:37	73.33%	53.85%	0.3798	1.4874	1.0000e-05
6	36	00:02:46	80.00%	57.69%	0.1493	1.5413	1.0000e-05
6	38	00:02:53	80.00%	59.62%	0.3355	1.5603	1.0000e-05
6	40	00:03:00	86.67%	59.62%	0.4197	1.5660	1.0000e-05
6	42	00:03:07	86.67%	59.62%	0.7130	1.5711	1.0000e-05
6	44	00:03:16	80.00%	61.54%	0.3920	1.5398	1.0000e-05
7	46	00:03:23	100.00%	61.54%	0.1493	1.5339	1.0000e-05

Training Data Over Time



Random Image Set with Prediction and Confidence

TEXT-TO-SPEECH

Once the network is trained and combined with the USB webcam, employing text-to-speech is relatively simple. I utilized a user-created text-to-speech function for this project, so all I needed to do was call it inside the continuous loop.

```
%%
camera = webcam('Logitech Webcam C930e');
i = 1;
while true
    picture = camera.snapshot;
    picture = imresize(picture,[227,227]);

    label = classify(alexTransfer, picture);

    image(picture);
    title(char(label));
    if i == 25
            tts(char(label));
            i = 1;
    end
    i = i + 1;
    drawnow;
end
```

CONCLUSIONS AND FUTURE WORK

- Higher accuracy in classifying objects and buildings can be achieved through more images in the dataset, tweaking with training options, or even implementing training on a GPU
- The eventual goal is to turn this MATLAB code into an application, primarily to be used with a smartphone for the greatest convenience and portability
- Group 2 images are not very similar to original ImageNet dataset, so training Group 2 images with the original ILSVRC images could be a viable option
- I'd like to expand our range to a wider area than just the SCU campus
- Could be integrated with GPS to not only identify objects, but calculate your precise location as well

An Input Power-Aware Efficiency Tracking Technique for Energy Harvesters in IoT

PhD Student: Sanad Kawar. Advisors: Shoba Krishnan and Khaldoon Abugharbieh
Department of Electrical and Computer Engineering

Santa Clara University
School of Engineering

Introduction

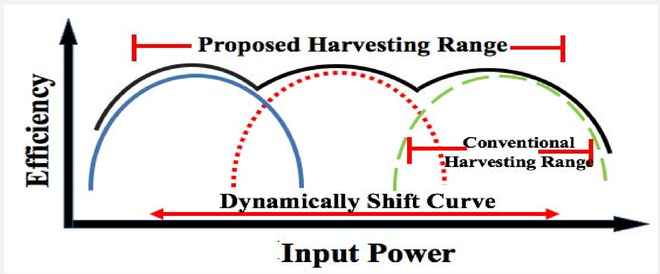
- **The Internet of Things (IoT):**
  - Enables intelligent monitoring and management
  - Small Wireless Sensor nodes (WSN) ➡ battery powered
  - Limited battery capacity ➡ periodic replacement costly/hard
- **Solution: Energy harvesting (EH)**
  - Recharge batteries or storage capacitors
  - Energy-autonomous, minimizes maintenance costs
  - Harvest from DC sources (light and thermal), AC (vibration) or RF

Challenges

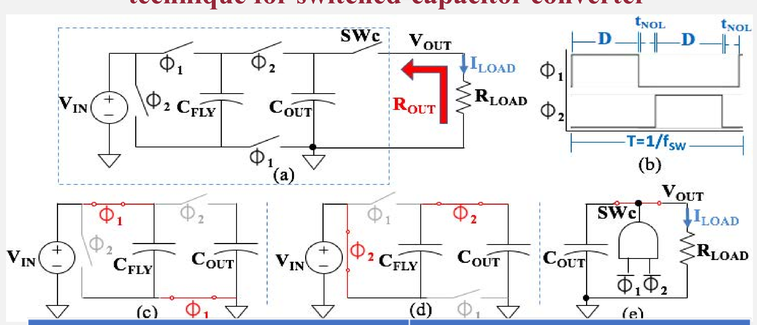
1. Harvested voltage too low ➡ need to boost voltage
   - Solar cell typical output ~0.3-0.6V
   - Too low to charge batteries or power CMOS
2. Environment-dependent energy
   - mm-scale solar cell: 20nW-200uW depending on illuminance
3. Low ambient power levels ➡ difficult to achieve high efficiency
   - Quiet, cold and/or dark environment: pW-nW in some applications

Objective

Extend input power harvesting range to accommodate low power IoT



Previous Work: Discontinuous Charging technique for switched-capacitor converter



Conventional (Continuous)	Proposed (Discontinuous)
Continuously provides current to load	Provides current to the load only during t_{NOL}
$I_{LOAD-AVG}=I_{LOAD-PEAK}=V_{OUT}/R_{LOAD}$	$I_{LOAD-AVG}=2.t_{NOL}I_{LOAD-PEAK}=2.t_{NOL}.V_{OUT}/R_{LOAD}$
t_{NOL} minimum(~1%) to achieve duty cycle ~50%	Up to 20% of periods

Achieving Max. Efficiency at Anticipated P_{IN}

Available Input Power Drops ➡ Reduce the switching frequency

Lower parasitic bottom plate and switching losses ➡ $P_{SW}+P_{BOT}=fswC_{GGtot}V_{IN}^{2}+0.1f_{sw}C_{FLY}V_{IN}^{2}$

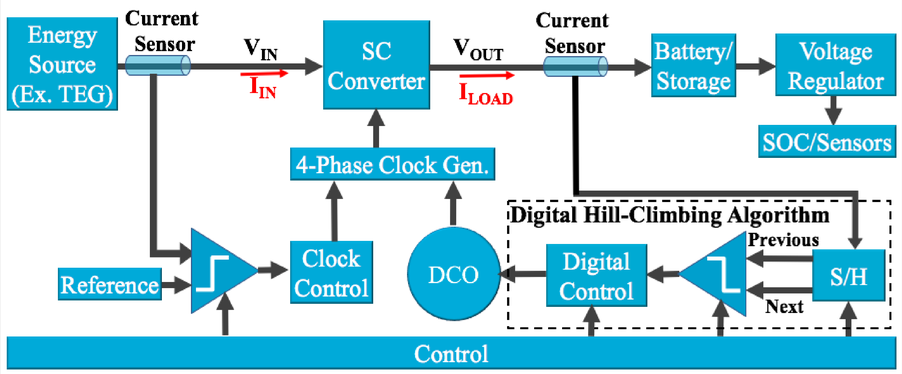
R_{OUT} α $(1/f_{SW})$ ➡ P_{ROUT} and V_{OUT} at lower f_{SW} would normally increase

$V_{OUT}=2V_{IN}-I_{LOAD-avg}R_{OUT}$ ➡ $P_{ROUT}=I^{2}_{LOAD-avg}R_{OUT}$

Supply current only during clock non-overlap ➡ Reduce t_{NOL} to reduce $I_{LOAD-AVG}$

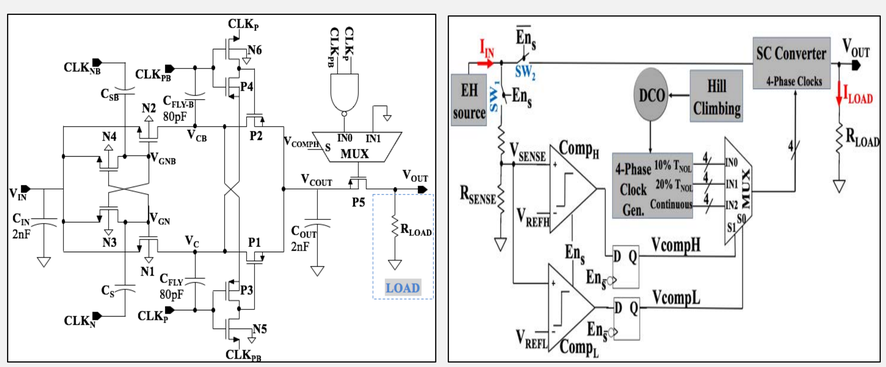
Converter maintains P_{ROUT} and V_{OUT} at higher R_{OUT} ➡ Converter achieves same power and voltage conversion efficiencies with less input power due to minimized losses.

Proposed System Architecture



- **Non-Overlap Time Selection:**
  - Senses input power, select continuous or discontinuous operation and non-overlap time
  - Shifts Entire Power Conversion Efficiency (PCE) curve
- **Switching Frequency Selection:** Hill-climbing algorithm to find maximum within curve

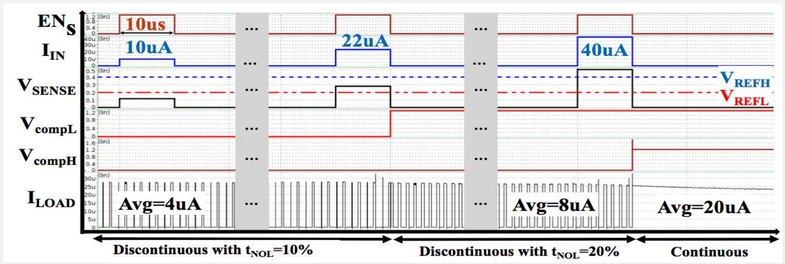
Circuits



Cross-Coupled SC Voltage Doubler with continuous/discontinuous charging

P_{IN} Sensing and Clock Control

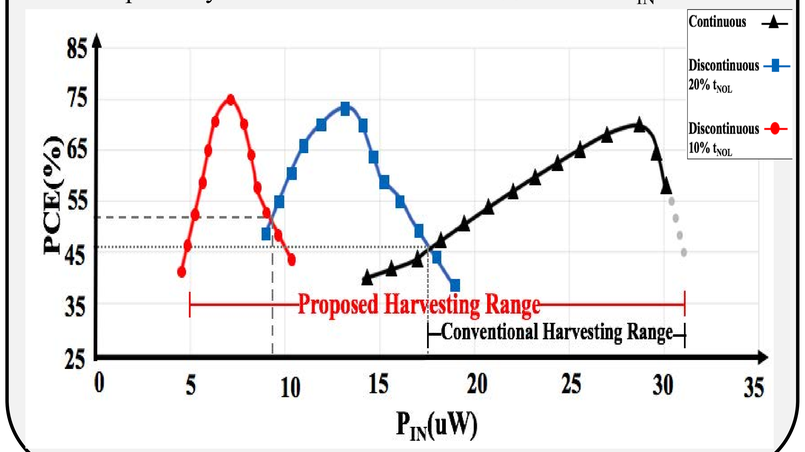
Simulation Results: P_{IN} Sensing and Clock Control



I_{IN} Range	V_{SENSE} Range	Operation Mode	t_{NOL}	$I_{LOAD-avg}$
I_{IN} < 20uA	$V_{SENSE}<V_{REFL}$	Discontinuous	10%	$0.2I_{LOAD-peak}$
20uA< I_{IN}<40uA	$V_{REFL}<V_{SENSE}<V_{REFH}$	Discontinuous	20%	$0.4I_{LOAD-peak}$
I_{IN} >40uA	$V_{SENSE}>V_{REFH}$	Continuous	<1%	$I_{LOAD-peak}$

Simulations Results: Power Conversion Efficiency

- Boosting 0.6V to 1.2V with a fixed 40kΩ load.
- Conventional/Continuous system's PCE < 45% if P_{IN} < 18uW
- Proposed system maintains PCE > 45% down to P_{IN} 5uW



Conclusion

- Compared to conventional MPPT, the proposed technique extends the harvesting range, by lowering the harvesting floor by 72%.
- PCE is optimized based on sense input power.

References

[1] X. Wu *et al.*, "A 20-pW Discontinuous Switched-Capacitor Energy Harvester for Smart Sensor Applications," *IEEE J. Solid-State Circuits*, vol. 52, no. 4, pp. 972–984, Apr. 2017.
[2] M. R. Elhebeary, M. A. A. Ibrahim, M. M. Aboudina, and A. N. Mohieldin, "Dual-Source Self-Start High-Efficiency Microscale Smart Energy Harvesting System for IoT," *IEEE Trans. Ind. Electron.*, vol. 65, no. 1, pp. 342–351, Jan. 2018.
[3] X. Liu, L. Huang, K. Ravichandran, and E. Sanchez-Sinencio, "A Highly Efficient Reconfigurable Charge Pump Energy Harvester With Wide Harvesting Range and Two-Dimensional MPPT for Internet of Things," *IEEE J. Solid-State Circuits*, vol. 51, no. 5, pp. 1302–1312, May 2016.
[4] T. Ozaki, T. Hirose, H. Asano, N. Kuroki, and M. Numa, "Fully-Integrated High-Conversion-Ratio Dual-Output Voltage Boost Converter With MPPT for Low-Voltage Energy Harvesting," *IEEE J. Solid-State Circuits*, vol. 51, no. 10, pp. 2398–2407, Oct. 2016.
[5] M. D. Seeman, "A Design Methodology for Switched-Capacitor DC-DC Converters," University of California, Berkeley,CA, UCB/EECS-2009-78, May 2009.
[6] S. Kawar, S. Krishnan, and K. Abugharbieh, "A Discontinuous Charging Technique with Programmable Duty-Cycle for Switched-Capacitor Based Energy Harvesting Circuits in IoT Applications," in *the 31st IEEE International System-On-Chip Conf.*, 2018, pp. 19–22.
[7] S. Kawar, S. Krishnan, and K. Abugharbieh, ÒAn Input Power-Aware Efficiency Tracking Technique for Energy Harvesters in IoT in Int. Symp. on Circuits and Systems , Sapporo, Japan, 2019

Distortion Analysis for a Spherical Transmitter

Amritpal Singh and Dr. Kurt Schab
Department of Electrical and Computer Engineering, Santa Clara University

Santa Clara University
School of Engineering

Background

- Linear time invariant (LTI) antennas face **limitations** that may be surpassed by nonlinear or time variant (nLTI) antennas.
- We would like to make a **quantitative comparison** between LTI and nLTI antenna systems.

LTI

- A system which forms a linear map between its inputs and outputs.
- A system for which a delay in the input causes the same delay in the output.
- The above properties make LTI systems easy to analyze.

Bandwidth

- The length of the range of frequencies present in a signal.
- Also used to describe the operational range of an antenna.
  - **Undefined for nLTI antennas**.

Objective

- Run circuit simulations of LTI and nLTI antenna systems and compute distortion for each.
- Use distortion values to compare the two systems.

Distortion

A measure of how different one signal is from another in the time domain.

$$d = \frac{1}{T}\int_0^T |\tilde{v}_\mathrm{o} - \tilde{v}_\mathrm{i}|^2 \,\mathrm{d}t$$

- Minimize over time delay.
- Normalize signals.
- Can be computed in either time domain or frequency domain using Parseval's Theorem.

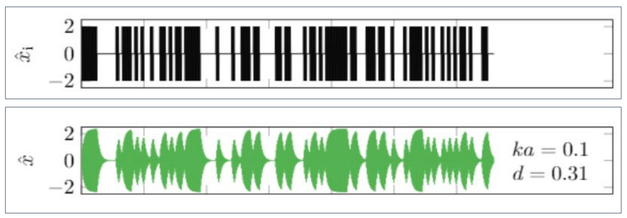


Figure 1. An on-off key (OOK) signal (black) along with a transmitted signal (green).

Method

Construct equivalent circuits for LTI and nLTI transmitters.

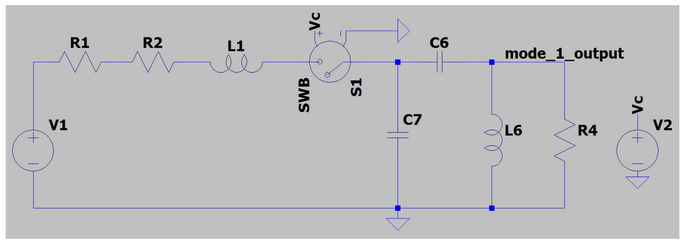


Figure 2. Equivalent circuit for a spherical transmitter. Here we have the nLTI variant, which includes a switch, denoted S1. The LTI system has the same circuit configuration except for this switch.

Run circuit simulation. LTspice®

Use generated data to compute distortion for both systems. MATLAB®

Update system parameter of interest (e.g. efficiency). Q, η, ϕ

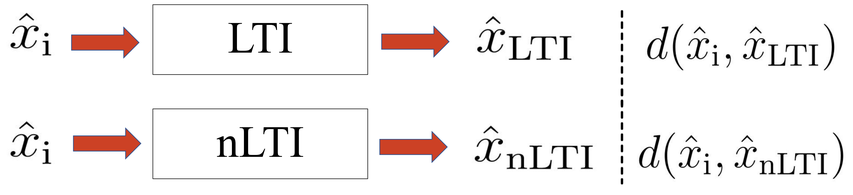


Figure 3. An ideal signal is passed into both the LTI and nLTI antenna systems. Both systems will then output signals in the time domain. The ideal and output signals are then used to compute distortion.

Spherical Transmitter

- The Chu sphere is a theoretically convenient, idealized antenna system which has the lowest possible Q factor. [2]
- In this work, we study signals using OOK modulation transmitted by both LTI and nLTI Chu spheres.

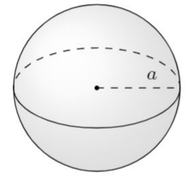
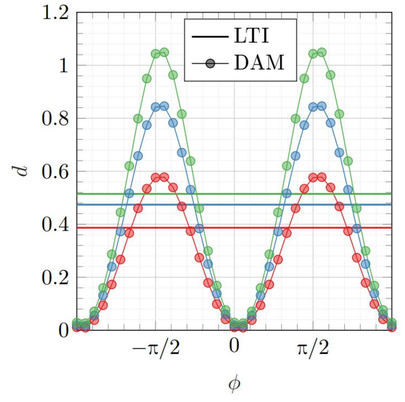


Figure 4. Spherical transmitter model. [1]

Application

An implementation of the DAM scheme would necessarily have non-idealities present in the switch. Figure 5 [3] shows the effects of an asynchronous switch. Here we can observe the points at which the non-idealities in the DAM scheme would result in worse performance than the LTI scheme.



- **Horizontal lines**: LTI distortion
- **Curves**: nLTI distortion
- **Different colors**: varying signal characteristics

Figure 5. Distortion as a function of φ, which denotes the switch phase offset. φ equal to zero corresponds to ideal conditions.

Conclusion

- nLTI antenna systems may offer a way to quantifiably achieve **better performance** than LTI antenna systems for the same task.
  - This is especially important for applications requiring electrical small antennas.
- **Future work** may be directed by developing **other metrics** for comparison, such as efficiency, as well as switching schemes to **transmit signals other than OOK**.

References

[1] K. Schab, A. Singh, N. Bohannon. "Distortion Analysis for the Assessment of LTI and non-LTI Transmitters." *IEEE Trans. on Antennas and Propagation*. In review.
[2] L. J. Chu, "Physical Limitations of Omnidirectional Antennas", *J. Appl. Phys.*, vol. 19, pp. 1163-1175, Dec. 1948.
[3] K. Schab, A. Singh, N. Bohannon, D. Huang, and J. J. Adams, "Quantifying non-ideal switch and timing properties of direct antenna modulation transmitters using distortion analysis," *Allerton Antenna Applications Symposium*, Monticello, IL, Sept. 2019.

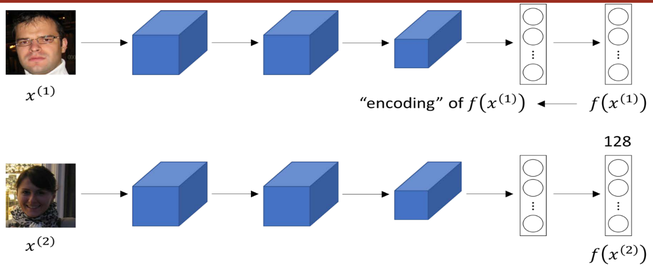
Face Recognition using Siamese Neural Network

Keshav Pawar[1], Prof. Tokunbo Ogunfunmi | Department of Electrical and Computer Engineering
Santa Clara University, 500 El Camino Real, Santa Clara, 95053

ABSTRACT

In today's world, security is a major issue. People are heavily using smartphones and are now starting to consider smart watches and even smart homes. Face recognition is a technique which can be used almost everywhere at present. For example: It can also be used in the Automobile industry for security purposes. The technique of identification of a person and allowing access can be carried out by a network called Siamese Neural Network. This network uses your image to identify who you are. Your unique identity is stored at the backend and then retrieved when the model finds a match of your image. Using this technique, we will check how this technique is applicable in various fields like Automobile, Real Estate, and E-commerce. Since this model uses face recognition, it is important to know how it can be used. The purpose of this research is to showcase some different fields in which the Siamese Neural Network would have great potential.

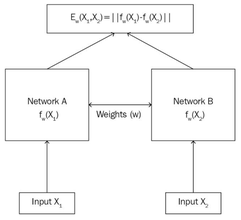
SIAMESE NEURAL NETWORK



https://www.kaggle.com/arpandhatt/siamese-neural-networks

Architecture

- The network uses same weights while working on two different vectors to generate comparable output
- This network uses One shot learning. That means, it trains quickly
- This network uses Triplet loss for accurate detection
- Triplet loss uses distances to compare between positive (truth input) and negative (false input)
- External memory to hold the results



http://aigradients.com/2019/06/19/siamese-networks-eli5/

APPLICATIONS

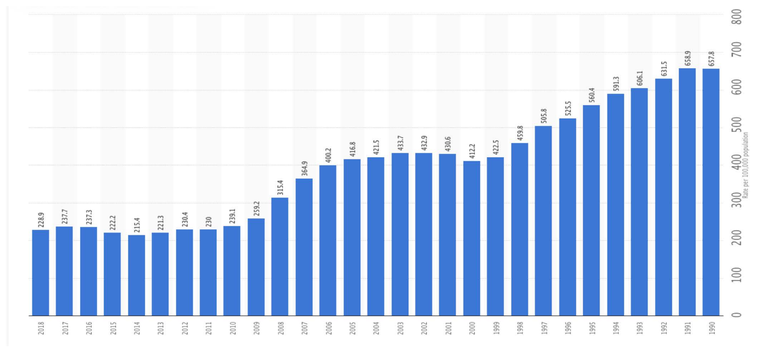
- Payment authentication in Smartphones
- Can be used in cameras in a Smart Home
- Automobile industry - To identify the owner/driver
- Can be an eliminator for biometric in some fields
- Authentication using a face detection device for Locker Rooms

EXISTING TECHNOLOGY

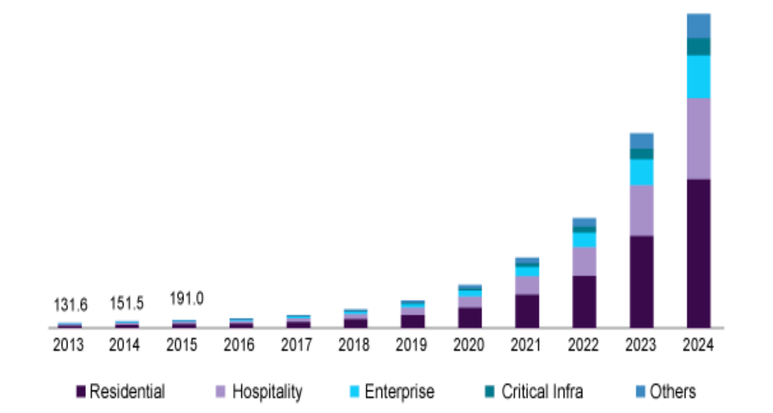
- Smart phones use facial id as a method to authenticate the payment process.
- Nest - Hello Camera for smart home face recognition.
- Embedded face recognition model in the car system. This is required because about $6 billion was lost in vehicle theft last year.

-https://visagetechnologies.com/face-recognition-in-cars/

MOTOR VEHICLE THEFTS IN UNITED STATES



SMART LOCK MARKET SIZE – UNITED STATES (USD MILLION)



https://www.grandviewresearch.com/industry-analysis/smart-lock-market

IMPLEMENTATION IN LOCKER SYSTEMS

- Can be used for authentication in Locker Systems
- Huge market for advancements in locker systems in United States
- One common face detection camera can be used to open the lockers
- A fast and effective way to connect the user identity (face) to the locker code
- The model can be easily trained on new faces for future users

PROS AND CONS OF USING SIAMESE NETWORK

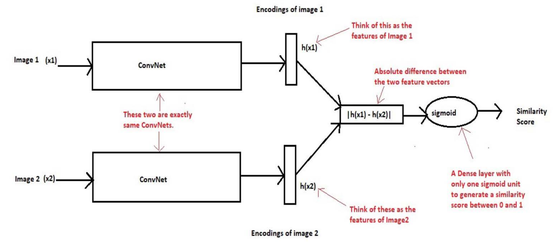
PROS:

- More robust to class imbalance - If the model has learned well what makes any 2 people the same, one example of a person will be enough to predict that person in the future. It does one shot learning!
- Nice to ensemble with the best classifier - Given that its learning mechanism is somewhat different from Classification, simple averaging it with a classifier can do much better than averaging 2 correlated supervised models

  .
- Better embeddings - Siamese focuses on learning embeddings that place same classes/concepts close together. Hence, it can learn semantic learning.

CON:

- Training involves Pairwise Learning which has quadratic pairs to learn from (in order to see all information available). This results in a slightly slower detection than Classification (pointwise learning)



https://towardsdatascience.com/one-shot-learning-with-siamese-networks-using-keras-17f34e75bb3d

POTENTIALS

- Integrating the face recognition model in the hardware to use in automobile/smart home/locker room cameras.
- Learning to train the model and add any user to it when required.
- Creating a secure and efficient environment around you.
- Integrating the Siamese Model can provide flexibility to quickly act on a recently added dataset.
- Implementing a design to ensure that your assets are safe.

CONCLUSIONS

- One shot learning is fast, accurate and convenient when you're working in an agile environment.
- Due to the architecture of Siamese network, it makes it easy for the user to add and train the model.
- It has direct applications in the automobile, smart home industry and locker system advancement technology.

A Fast 2-D Convolution Technique for Deep Neural Networks

Anaam Ansari[1], Prof. Tokunbo Ogunfunmi[2]

[1,2] Signal Processing Research Lab, Santa Clara University

[1] aaansari@scu.edu,[2] togunfunmi@scu.edu

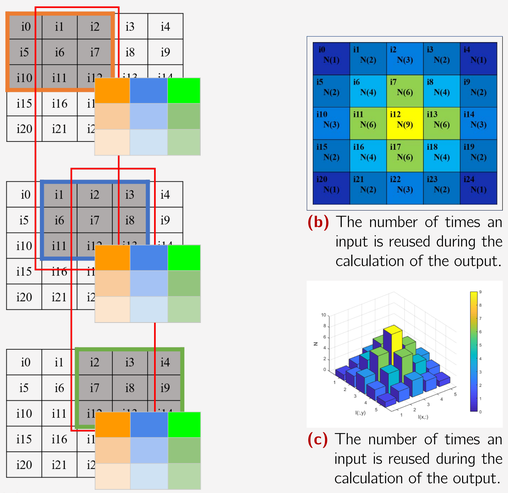
1 Abstract and Motivation

We introduce a new and improved 2-D convolution method called Single Partial Product 2-D Convolution (SPP2D Convolution) that will help calculate 2-D convolution in a fast and expedient manner. We demonstrate that the new SPP2D convolution will prevent recalculation of partial weights and we present theoretical analysis of our technique compared to some other popular techniques. In our analysis, we determine that our technique can reduce the clock cycles related to input reuse by at least $2.9\times$ times in comparison with the technique adopted in the work done in [1] and about $8\times$ than the standard sliding window approach.

2 Previous Work

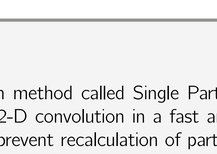
There are several methods to perform the convolution operation. Convolutions can be implemented using matrix multiplication as well as vector multiplication [2] however convolution using matrix multiplication introduces redundant operations by converting the input matrix into a Toeplitz matrix and convolutions using vector multiplication take a long time if done serially or require large memory transfers if the operation is carried out in parallel. Solutions like [3] propose reuse of input weights in order to save fetching them from off chip memory. Reuse is an important mechanism in CNNs as the traditional sliding window method without any reuse would result in fetching some input pixels and weights multiple times. In [4], the authors perform a redundancy analysis of the weights in order to not repeat sending similar data from off chip to on chip. In [1], the authors try to implement an architecture that aims at reducing the latency of networks by implementing the reuse of input pixels.

3 Single Partial Product 2-D Convolution



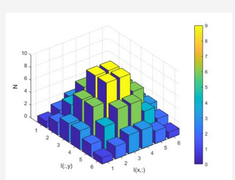
(a) Inputs reused during convolution computation

(b) The number of times an input is reused during the calculation of the output.

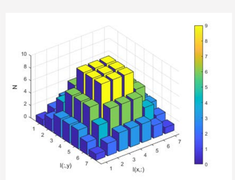


(c) The number of times an input is reused during the calculation of the output.

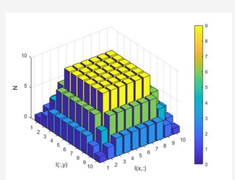
Figure 1: Conventional Sliding Window Convolution



(a) 6 × 6 input



(b) 7 × 7 input



(c) 10 × 10 input

Figure 2: Pattern of the frequency with which input pixels are needed in a sliding window operation

Table 1: Number of inputs with the frequency N(9),N(6),N(3),N(4),N(2),N(1) for various input sizes

input size	N(9)	N(6)	N(3)	N(4)	N(1)	N(2)
5	1	4	4	4	4	8
6	4	8	8	4	4	8
7	9	12	12	4	4	8
10	36	24	24	4	4	8
14	100	40	40	4	4	8
28	576	96	96	4	4	8
56	2704	208	208	4	4	8
112	11664	432	432	4	4	8
224	48400	880	880	4	4	8

Table 2: Frequency calculation based on input dimensions

No. of inputs with N(9)	$(H_{input}-2)\times(W_{input}-2)$
No. of inputs with N(6) and N(3)	$(H_{input}-4)*2+(W_{input}-4)*2$
No. of inputs with N(4) and N(1)	4
No. of inputs with N(2)	$2*4$

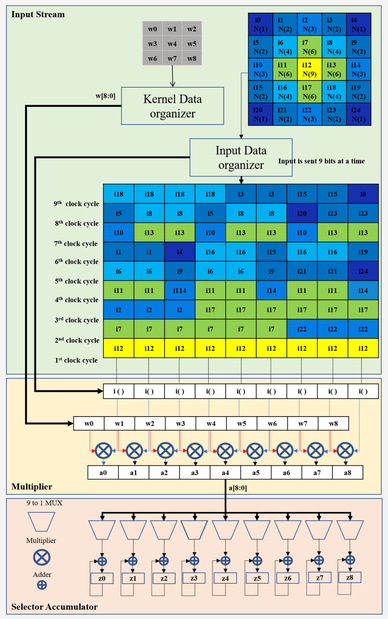
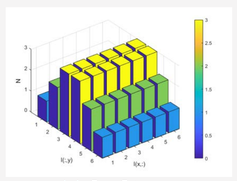


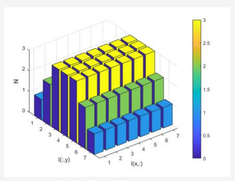
Figure 3: Architecture to perform SPP2D Convolution

Table 3: Multiplying inputs with weights in order in which they are most used. N(x) gives the number of times an input pixel used in generating the outputs

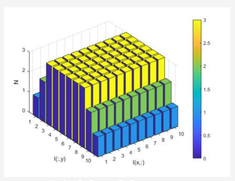
clock cycles	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
	N(9)	N(6)	N(6)	N(6)	N(6)	N(4)	N(4)	N(4)	N(4)	N(3)	N(3)	N(3)	N(3)	N(2)	N(2)	N(2)	N(2)	N(2)	N(2)	N(2)	N(2)	N(1)	N(1)	N(1)	N(1)
	i12	i7	i11	i17	i13	i2	i14	i22	i10	i6	i8	i18	i16	i1	i13	i5	i9	i15	i19	i21	i23	i0	i4	i20	i24
w0	w0i12	w0i7	w0i11			w0i2			w0i10	w0i6				w0i1		w0i5						w0i0			
w1	w1i12	w1i7	w1i11		w1i13	w1i2				w1i6	w1i8			w1i1	w1i3										
w2	w2i12	w2i7			w2i13	w2i2	w2i14				w2i8			w2i3		w2i9						w2i4			
w3	w3i12	w3i7	w3i11	w3i17					w3i10	w3i6			w3i16			w3i5		w3i15							
w4	w4i12	w4i7	w4i11	w4i17	w4i13					w4i6	w4i8	w4i18	w4i16												
w5	w5i12	w5i7		w5i17	w5i13		w5i14				w5i8	w5i18					w5i9		w5i19						
w6	w6i12		w6i11	w6i17				w6i22	w6i10				w6i16					w6i15		w6i21			w6i20		
w7	w7i12		w7i11	w7i17	w7i13			w7i22				w7i18	w7i16							w7i21	w7i23				
w8	w8i12			w8i17	w8i13		w8i14	w8i22				w8i18							w8i19		w8i23				w8i24



(a) 6 × 6 input



(b) 7 × 7 input



(c) 10 × 10 input

Figure 4: Pattern of the frequency with which input pixels are needed in [1]

Table 4: Optimized input stream

clock cycles	1	2	3	4	5	6	7	8	9
	i12 N(9)	**i17 N(6) i22 N(3)**	**i17 N(6) i2 N(3)**	**i11 N(6) i14,N(3)**	**i6 N(4) 19 N(2) i21 N(2) i24 N(1)**	**i16 N(4) i1 N(2) i19 N(2) i4 N(1)**	**i10 N(6) i13 N(3)**	**i8 N(4) i5 N(2) i23 N(2) i20 N(1)**	**i18 N(4) i3, N(2) i15, N(2) i0 N(1)**
w0	w0i12	w0i7	w0i2	w0i11	w0i6	w0i1	w0i10	w0i5	w0i0
w1	w1i12	w1i7	w1i2	w1i11	w1i6	w1i1	w1i13	w1i8	w1i3
w2	w2i12	w2i7	w2i2	w2i14	w2i9	w2i4	w2i13	w2i8	w2i3
w3	w3i12	w3i7	w3i17	w3ii11	w3i6	w3i16	w3i10	w3i5	w3i15
w4	w4i12	w4i7	w4i17	w4i11	w4i6	w4i16	w4i13	w4i8	w4i18
w5	w5i12	w5i7	w5i17	w5i14	w5i9	w5i19	w5i13	w5i8	w5i18
w6	w6i12	w6i22	w6i17	w6i11	w6i21	w6i16	w6i10	w6i20	w6i15
w7	w7i12	w7i22	w7i17	w7i11	w7i21	w7i16	w7i13	w7i23	w7i18
w8	w8i12	w8i22	w8i17	w8i14	w8i24	w8i19	w8i13	w8i23	w8i18

4 Results

We can formulate a way to calculate the number of clock cycles it takes to perform an SPP2D convolution given an input of size $H_{input} \times W_{input}$ based on our understanding of the complementary sets. No. of clock cycles for SPP2D: No of inputs with N(9)+ $\frac{1}{2}\times$(No. of inputs with N(6) + No of inputs with N(3)) + $\frac{1}{4}\times$ (No. of inputs with N(4) + No of inputs with N(2) + No of inputs with N(1)). This can be summarized as follows :

$$\text{SPPD clock cycles} = H_{input} * W_{input} + (H_{input} - 4) * 4 + 4 \quad (1)$$

For example, we have demonstrated that the number of clock cycles to perform convolution on 5×5 with 3×3 takes 9 clock cycles to generate an output of size 3×3 and it takes 17 clock cycles if the output is of size 5×5 . This can be scaled for inputs of higher dimensions such as the input dimensions of the layers of VGGNet16 network. The results are described in Table 5. Columns *a, b, c* indicate the theoretical number of clock cycles needed to perform a convolution operation on an input of a certain dimension using the sliding window, [1], and SPP2D convolution **(w/o padding)** and columns *d, e, f* indicate the theoretical number of clock cycles need by sliding window, [1], and SPP2D convolution **(with padding)**. From the table we can see that the SPP2D convolution takes about $2.9\times$ (columns (b/c, e/f) less clock cycles than [1] and $8\times$ (columns (a/c), (d/f)) less clock cycles than the sliding window technique for both operation types - w/o padding and with padding.

Table 5: Theoretical calculation of the number of clock cycles needed for VGG16 layers inputs

	without padding							**with padding**						
	a	b	c	a-c	a/c	b-c	b/c	d	e	f	d-f	d/f	e-f	e/f
input size	Sliding Window (w/o padding)	[1] (w/o padding)	SP2D Conv w/o padding					Sliding Window (w padding)	[1] (w padding)	SP2D Conv (w padding)				
5	81	45	17	64	4.76	28	2.65	225	105	49	176	4.59	56	2.14
6	144	72	28	116	5.14	44	2.57	324	144	64	260	5.06	80	2.25
7	225	105	41	184	5.49	64	2.56	441	189	81	360	5.44	108	2.33
10	576	240	92	484	6.26	148	2.61	900	360	144	756	6.25	216	2.50
14	1296	504	188	1108	6.89	316	2.68	1764	672	256	1508	6.89	416	2.63
28	6084	2184	776	5308	7.84	1408	2.81	7056	2520	900	6156	7.84	1620	2.80
56	26244	9072	3128	23116	8.39	5944	2.90	28224	9744	3364	24860	8.39	6380	2.90
112	108900	36960	12536	96364	8.69	24424	2.95	112896	38304	12996	99900	8.69	25308	2.95
224	443556	149184	50168	393388	8.84	99016	2.97	451584	151872	51076	400508	8.84	100796	2.97

5 Conclusion

SPP2D Convolution technique has theoretically proven to be a faster convolution technique and can help reduce latency of deep neural networks. It not only reduces the number of clock cycles that are needed to arrive at the output but also avoids recalculating the partial products. The proof of concept of SPP2D Convolution is very promising and can be adopted to accelerate deep neural network hardware. We are currently implementing in hardware an architecture that aims to be network agnostic and this method will help us implement convolutions of various sizes in a fast and efficient manner.

References

[1] A. Ardakani, C. Condo, M. Ahmadi, and W. J. Gross, "An architecture to accelerate convolution in deep neural networks," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 65, no. 4, pp. 1349–1362, 2017.

[2] V. Sze, Y.-H. Chen, T.-J. Yang, and J. S. Emer, "Efficient processing of deep neural networks: A tutorial and survey," *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2295–2329, 2017.

[3] Y. H. Chen, J. Emer, and V. Sze, "Eyeriss: A spatial architecture for energy-efficient dataflow for convolutional neural networks," *IEEE Micro*, pp. 1–1, 2017.

[4] A. Ansari and T. Ogunfunmi, "Selective data transfer from DRAMS for CNNs," in *2018 IEEE International Workshop on Signal Processing Systems (SiPS)*, Oct 2018, pp. 1–6.

Reconfigurable Substrate Integrated Waveguides Using Liquid Metal Microvasculature

Bradley Shirley[1], Marlee Strong[2], Dr. Jason Patrick[2], Dr. Kurt Schab[1]
[1]Department of Electrical and Computer Engineering, Santa Clara University
[2]Department of Civil Engineering, North Carolina State Univeristy

Motivation / Objectives

- Many antenna and microwave systems are built for single use applications
  - **Reconfigurable** and **multi-use** systems are achievable with liquid metal
  - From **frequency response, radiation pattern,** or **polarization**
- Use **microvascular** channels to route **liquid metal** to achieve a preferred state
- Reduce **cost**, mitigate **weight requirements**, and allow for more **compact designs**

Background

- Traditional waveguides guide electromagnetic waves with **minimal loss** to a desired location
- A new form of waveguide called a **Substrate Integrated Waveguide (SIW)** is implemented in this design
- Working in conjunction with the multifunctional composites group at North Carolina State University in Raleigh NC, we are able to fabricate and test these novel structures

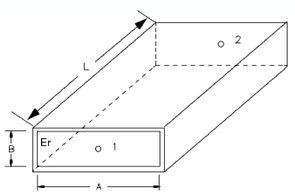


Figure 1: Standard Rectangular Waveguide [1]

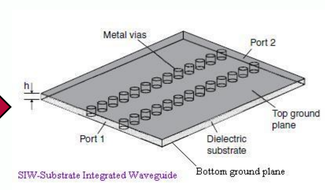


Figure 2: Substrate Integrated Waveguide [2]

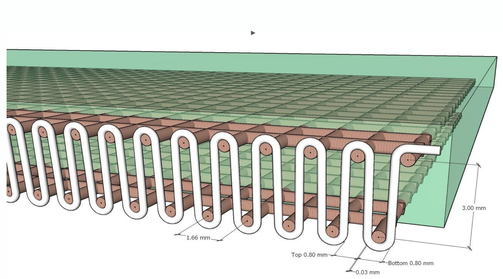


Figure 3: Implementation of microvascular channels

Research Plan

Design control SIW to obtain characteristics

Figure 5: From left to right: TRL calibration standards (through, reflect, line) and example device

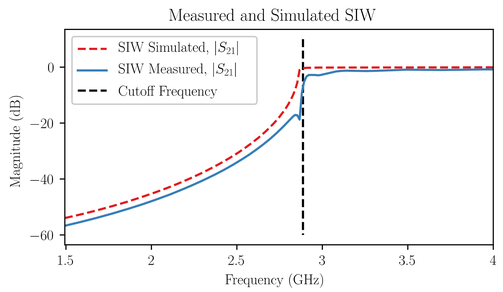


Figure 6: Measured Fabricated DUT data

Verify properties of liquid metal SIW

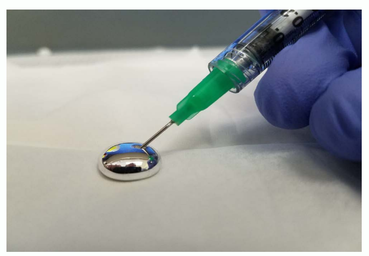


Figure 7: Eutectic Gallium-Indium Liquid Metal

Develop reconfigurable liquid metal SIW based devices

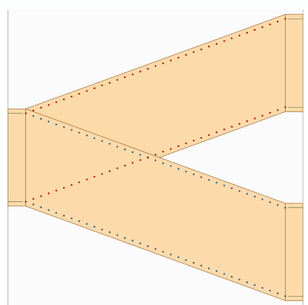


Figure 8: SIW Reconfigurable Device

Future Work

- **Measure** liquid metal implemented composites from North Carolina team
  - **Experimentally verify** liquid metal SIW transmission line properties
  - Co-optimize electrical and structural device characteristics
- Add **multiple channels** to SIW to allow for reconfigurability within one system
  - Create a **"railroad" switch** to divert energy to multiple locations using different reconfigurable states
  - Create a **variable attenuator** to allow for various states of attenuation depending on reconfigurable states
- Continue with **further applications** of **reconfigurability** using liquid metal

Conclusion

- Using liquid metal as a method of reconfigurability within larger communication systems is a viable option
- Having reconfigurable microwave and antenna systems allows for robust applications using minimal resources
- This application is one of many possibilities that can be achieved using these methods

References

[1] "RWG (Rectangular Waveguide)", *keysight.com.* Online. Available: http://edadocs.software.keysight.com/pages/viewpage.action?pageId=5928541 [Accessed May 28, 2019].

[2] "Substrate Integrated Waveguides", *microwaves101.com.* Online. Available: https://www.microwaves101.com/encyclopedias/substrate-integrated-waveguide. [Accessed May 28, 2019].

Proof of Concept of Swarm Scalar Field Adaptive Navigation

Shae T. Hart
Advisor: Dr. Christopher A. Kitts

Abstract

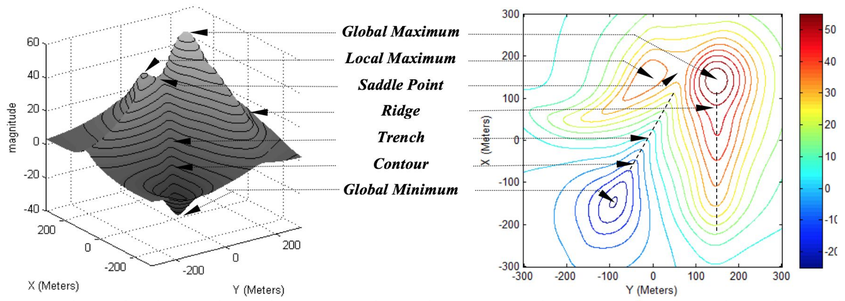
The Robotic Systems Laboratory researchers have previously developed a complete set of cluster-based adaptive navigation behaviors for locating or moving along all critical features of a scalar field. Our new work seeks to replicate this capability using a decentralized swarm control architecture that does not require full degree of freedom formation control. Swarm control strategies allow for decentralized control of many simple robots to perform collective behaviors based on local interactions.

Background

- *Centralized vs Decentralized Control:* In centralized control, control decisions use global information and are typically computed at a central location. In decentralized control, each robot makes its own control decisions based on local information from a limited number of neighboring robots.
- *Swarm Robotics* describes systems of large numbers of simple robots. The swarms used in this research are assumed to be homogeneous and use a decentralized control architecture to control each robot's velocity.
- *Current Working Swarm Definition:*
  - Constant velocity robots
  - Limited communication range
  - Reactive control system that doesn't use previous swarm state information

Swarm Scalar Field Adaptive Navigation

Traditional scalar field navigation required the entire field of interest to be mapped before robots could navigate to features of interest. Scalar field adaptive navigation allows robots to dynamically navigate to features of interest in the scalar field.

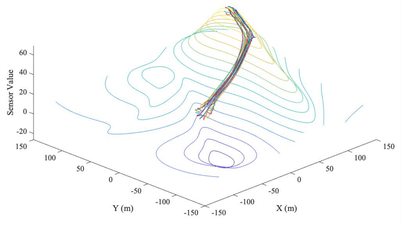


Hypothesis: Scalar field adaptive navigation can be performed using decentralized control of robotic swarms.

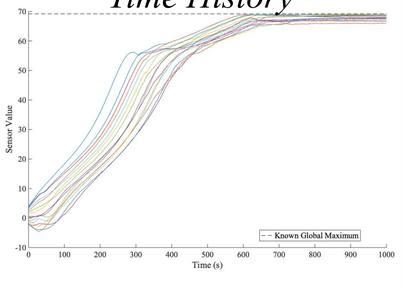
Simulation Results

Maxima Seeking:
Seeking maximum of scalar field

World View

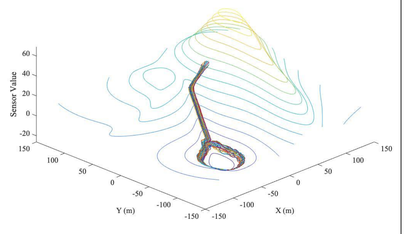


Time History

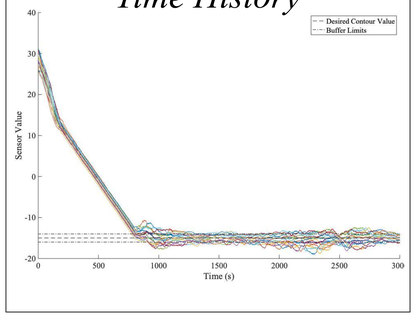


Contour Following:
Find and follow specified scalar value

World View

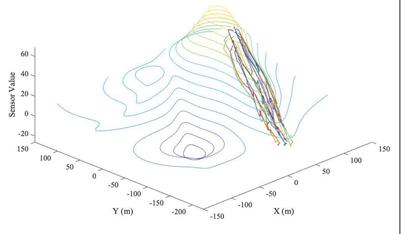


Time History

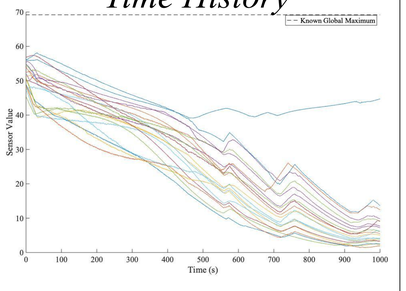


Ridge Following:
Follow minimum slope down from maxima

World View

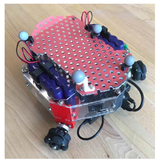


Time History



Experimental Testbed

The RSL's Indoor Testbed uses up to 10 holonomic omnibots. We continue to test iterations of composite behaviors using the testbed to ensure behaviors work when subjected to real-world conditions like noise.



Conclusion

Using swarm robotics for adaptive navigation provides a novel approach to adaptive navigation problems. This work shows that it is possible to use decentralized control of swarm robots to perform certain scalar field adaptive navigation behaviors.

Future Work

- Proof of concept of saddle point identification
- Unify and formalize control architecture for all adaptive navigation behaviors
- Clarify definition of swarm to include heuristics on how to define swarm parameters
- Stability analysis via Monte Carlo Simulation

Acknowledgements

SCU School of Engineering
The Robotic Systems Laboratory
Dr. Michael Neumann & Robert McDonald
Nathan Metzger & Max Reese

Exploring 3-Dimensional Environmental Phenomena with Drone Clusters

Robert K. Lee
Advisors: Dr. Christopher Kitts and Dr. Michael Neumann

Santa Clara University
School of Engineering

Abstract

Adaptive Navigation (AN) techniques have been developed to allow clusters of unmanned aerial vehicles to adaptively locate and move along features of interest in 3-dimensional environmental scalar fields, even if the field evolves over time. This allows us to efficiently perform tasks such as locating sources of pollution or radiation, characterizing the distribution of gas leaks and oil discharges, and following plumes to either their source or their impact zones. This work has been verified using high fidelity simulations of multi-drone clusters.

Applications

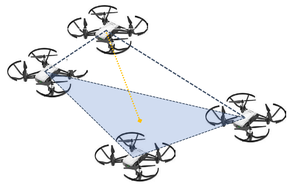
- **Environmental Sensing** – Locating origins and mapping extent of environmental signals: physical parameters, biological productivity, chemical dispersal
- **Disaster Response** – Finding gas or radiation leaks, setting up safety perimeters, characterizing impact regions and safe zones
- **Homeland Security** – Locating threats, finding low exposure approach paths, identifying extent of service areas, etc.

Acknowledgements

Robert McDonald
Robotic Systems Laboratory

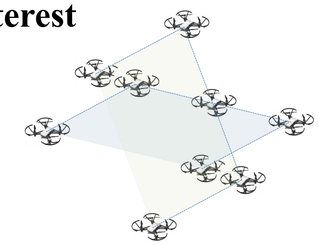
Summary

Vehicles move as part of reconfigurable formations called "clusters"
Vehicles sample the field's scalar value at their locations and share information
Local field characteristics are computed
Formation moves/morphs to navigate to/along features of interest



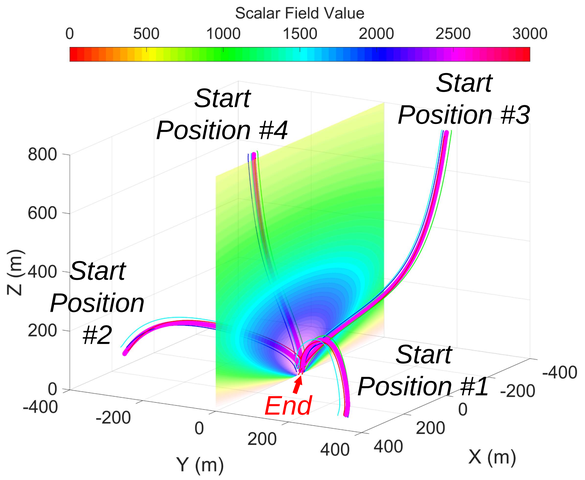
A 4-vehicle pyramid formation computes the 3-D field gradient, which is used for finding extreme points in a field and moving to/on iso-surfaces.

A 9-vehicle formation computes differential measurements within the field, aligns and straddles plumes, and navigates "down" the plume to the impact zone.



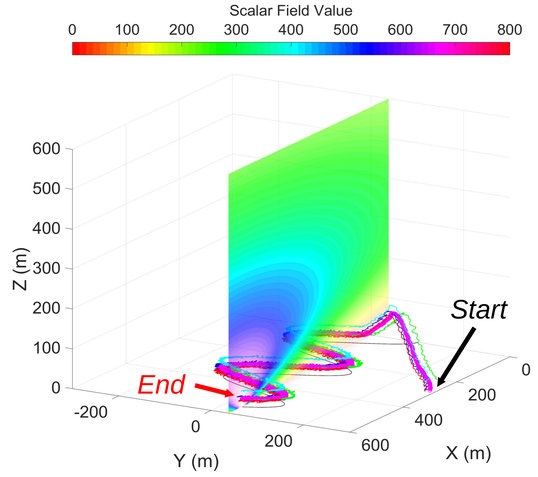
Locate a Stationary Source

Sample distributed field data, estimate field gradient, move cluster as a virtual rigid body in direction of the gradient.



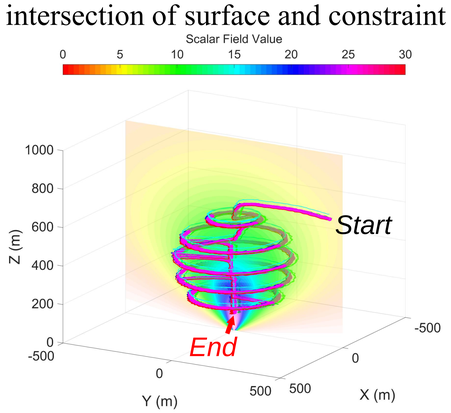
Locate a Moving Source

Source seeking strategy works for dynamic fields as long as the vehicles are faster than motion of the source.



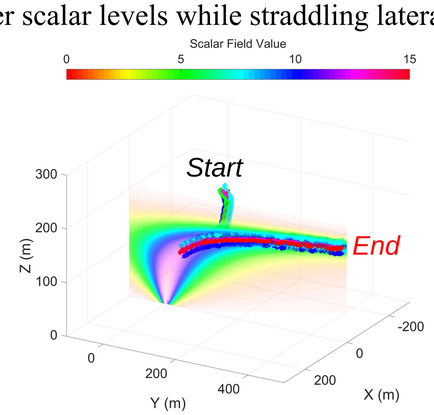
Map an Exposure Level Surface

Follow gradient to level of interest, follow level contour created by intersection of surface and constraint plane.

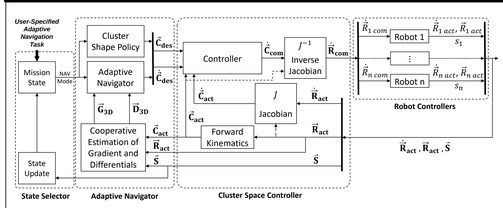


Follow a Plume to its Impact Region

Acquire plume, straddle plume, move longitudinally to lower scalar levels while straddling lateral peak.



Control Architecture



- **Robot Controllers**: Ensure platform stability and execute velocity commands
- **Cluster-Space Controller**: Maintains virtual structure attitude and geometry
- **Adaptive Navigator**: Steers and reconfigures the cluster based on the selected AN control primitive, such as gradient following, constrained contour tracking, etc.
- **State Selector:** Sequences AN primitives in order to implement mission-level capabilities such as spatially characterizing a scalar field within a region.

Conclusion

This work extends multirobot AN functionality, enabling navigation with respect to three dimensional environmental features. Specific contributions include introduction of new 3-D AN primitives, state-based sequencing to implement mission-level capabilities, and the use of high-fidelity simulation to verify the performance and capabilities of these techniques.

Aerodynamics of a Forward-Facing Fin

Tioga Benner and Godfrey Mungal

Department of Mechanical Engineering

Santa Clara University
School of Engineering

Rationale

Surgeon Fish

Observed tail spines of surgeonfish

Present in many different species, evolutionarily advantageous

Possible aerodynamic benefits

Conjecture

Forward fin creates reverse vortex with beneficial aerodynamic properties

Similar to properties found in humpback whale tubercles

Bio-inspired Design

Growing field of engineering design

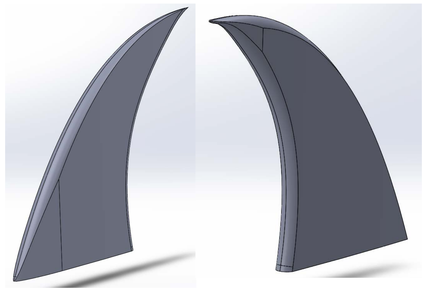
Designs inspired by natural world and evolution

Has resulted in innovative and sometimes groundbreaking products and discoveries

Solidworks Model

7 in chord length, 9.1 in span giving an aspect ratio of around 2

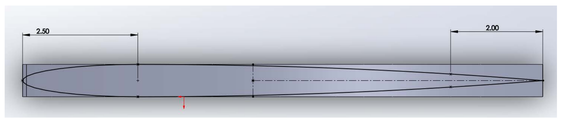
Both pointed and curved airfoil shapes



Forward and backward facing as well as straight fins

Propagating complex shape along curved path while changing size proved very challenging

Resulted in less aerodynamic fin than most airfoils



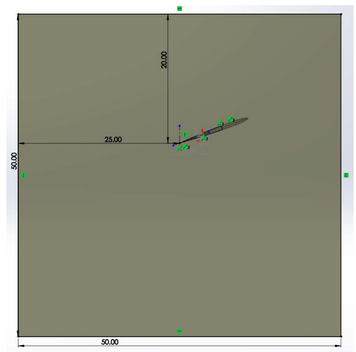
Simulated Results Using Star CCM+

Setup

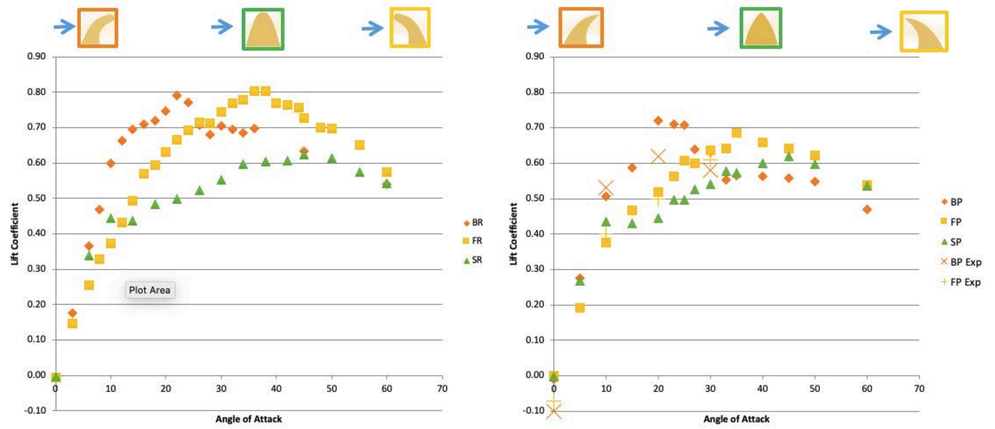
Fins placed in 50 by 50 by 25 inch box,

Water at 1 m/s giving Reynolds number of $2 * 10^5$

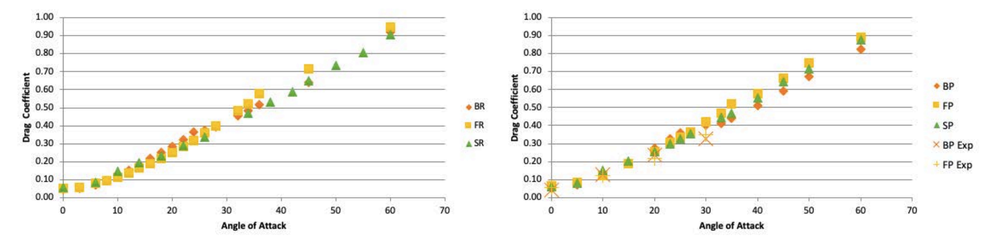
Segregated flow model with K-Omega Turbulence Model



Lift Coefficient



Drag Coefficient

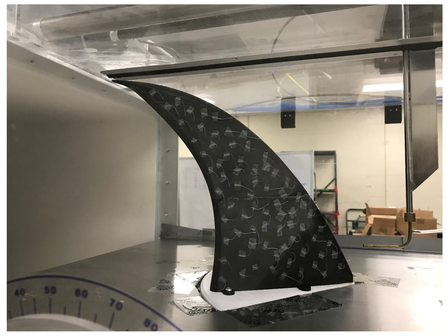


Wind Tunnel Setup

Used Santa Clara Wind Tunnel with 24 by 24 by 48 inch test section, Vmax ~50m/s

Tufts attached to fin for flow visualization but removed for actual resting with later surface visualization done using rubbing alcohol and chalk dust

Baseplate attached to fin to control angle and ground plane used to prevent baseplate from changing drag or lift measurements



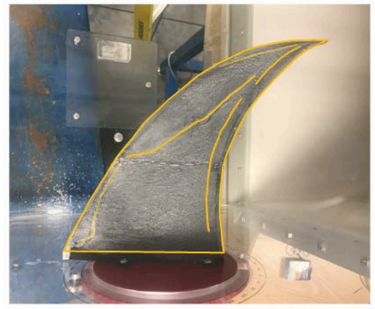
Surface Flow

Chalk dust and rubbing alcohol mixture used while wind tunnel is running, chalk accumulation line patterns were formed on the surface

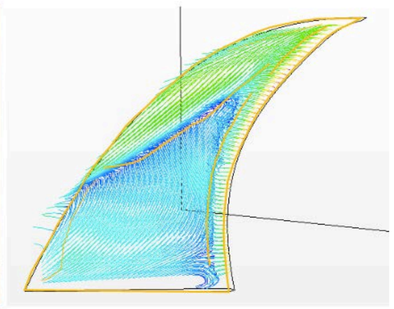
These lines were overlaid on the surface flow pattern found in Star CCM+

Showed remarkable agreement between the two cases

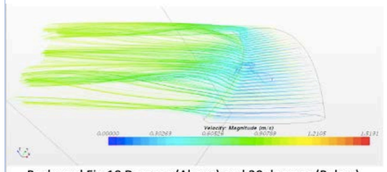
Experiment



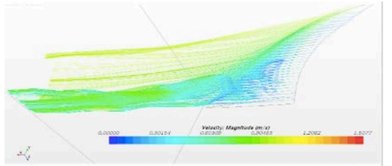
K-omega model



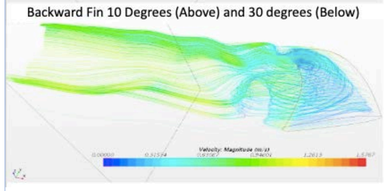
Streaklines

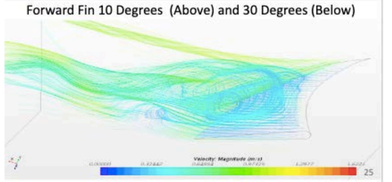


Backward Fin 10 Degrees (Above) and 30 degrees (Below)



Forward Fin 10 Degrees (Above) and 30 Degrees (Below)





Conclusions

Three different types of stall observed: Backward fin shows traditional sudden steep stall, Forward fin has a smoother rolling stall, Straight fin has two-time stall

Forward fins are advantageous when used at higher angles of attack, and reduce chance of extreme drops in lift resulting in increased safety

Similar results seen in tandem orientations

Shape Memory Alloy Research for Actuation in Extreme Environments

Rachel Stolzman[1], Ann McGuire[1], Dr. Geoff Wheat[2], Dr. Christopher Kitts[1]
[1]Santa Clara University; [2]University of Alaska, Fairbanks

ABSTRACT

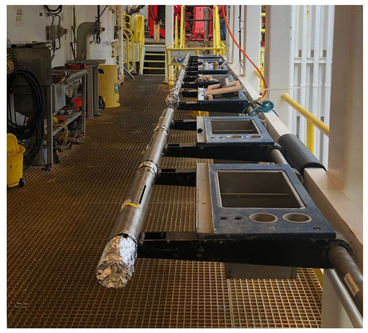
A titanium **sampler** is designed to extract and hold water from a **deep sea borehole** to **observe if life can exist in extremely high temperatures** thousands of meters below sea level. To actuate these samplers, several "trigger" mechanisms using **shape-memory alloys** (SMAs) were created to successfully capture water at various temperatures via thermal energy. SMAs **eliminate the need for electronics**, which would not be able to perform in such extreme environments.

BACKGROUND

Deep-Sea Boreholes: left behind from oil drilling and can reach depths of 5000 meters below the seafloor and temperatures up to 200℃

Integrated Ocean Discovery Program (IODP): international lead researching organization of boreholes responsible for verifying that designed sampler meets all borehole requirements and ship interfaces

IODP Expedition 385T: samplers were deployed for this expedition off the coast of Chile in late August, 2019



SPECIFICATIONS

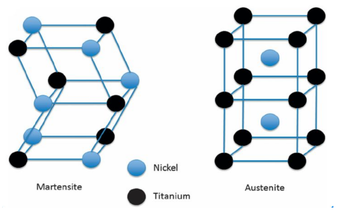
Borehole Dimensions	4 in. diameter
Sampler Target Diameter	3.75 in. diameter maximum
Temperature Range	22℃ - 200℃
Dynamic Load	~3000 lbs
Material	all components coming in contact with water sample must be anti-bacterial

SHAPE-MEMORY ALLOYS

Shape Memory Effect: materials "remember" a predetermined shape, are deformed, then return to original shape upon internal/external heating

Crystal structure: martensite at low temperatures, **austenite** at high temperatures

Common Metals: zinc, copper, aluminum, **nickel**, and **titanium**



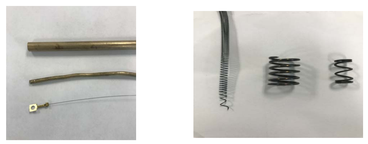
Crystal Structure of a NiTi SMA[1]

Temperature Dependence:

Heating: completely returns to original shape at the austenite finish temperature

Cooling: no deformation takes place

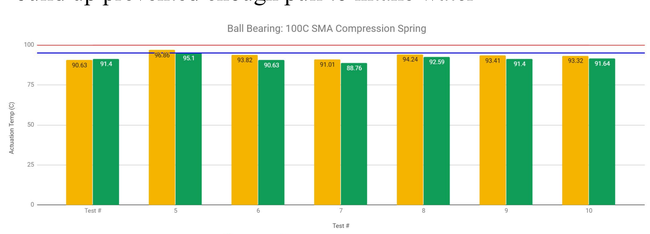
Shape Dependence: generally speaking, **only large deformations or large forces** can be obtained from an SMA, but not both



Left: 0.197in, 0.079in, and 0.006in Wire
Right: Tension, High Temperature Compression, and NiTi Compression Springs

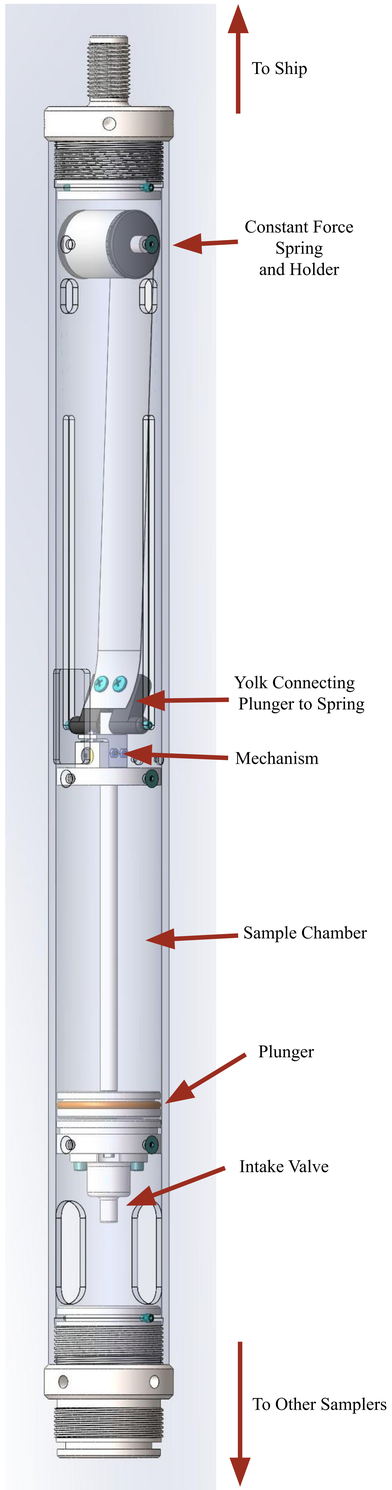
TESTING

- Water heated in tank to or beyond activation temperature of **ball bearing SMAs**
- Test terminated when the SMA completed **triggered water intake** or after an extended period of time of no movement beyond its activation temperature
- Common Problem: friction of constant force spring and water build up prevented enough pull to intake water



Results from 100℃ ball bearing tests in water tank

SAMPLER ASSEMBLY

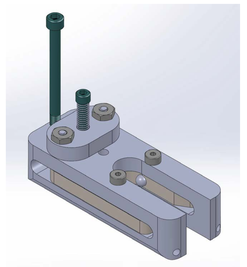


MECHANISMS

Ball Bearing Design (80℃-100℃)

SMA: compression or tension springs

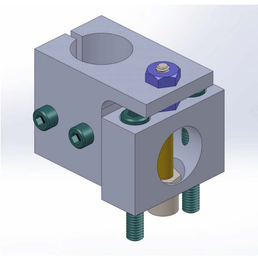
- Two ball bearings sit in the groove at the top of the plunger rod.
- Two pins hold the flippers in place by a stainless steel compression spring.
- Pins' **plate is lifted by the SMA spring** and the force from the constant force spring pulls up on the ball bearings, forcing the flippers out of the way.



Frangibolt Design (140℃-170℃)

SMA: frangibolts

- Custom blend of frangibolts
- The frangibolt sits in a retaining block
- An L-shaped bracket gets attached to the rod
- The **fastener breaks** and the plunger can move up the chamber



FUTURE WORK

Gas Tight Sampler: to measure dissolved gases and to prevent shearing of microbes

Reversible Gear Train Mechanism: open and close valve using continuous rotational motion

Updated Integration of SMAs: additional new ways of actuation

Additional IODP Expeditions

ACKNOWLEDGEMENTS

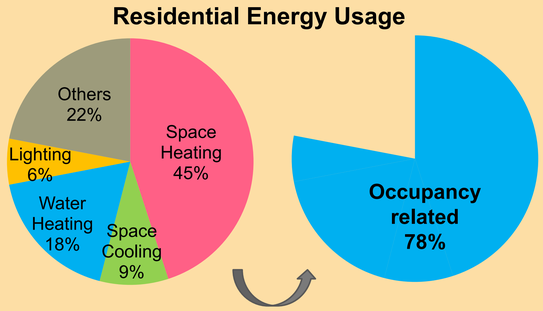
SCU Robotic Systems Laboratory
Monterey Bay Aquarium Research Institute
International Ocean Discovery Program
National Science Foundation
Anne Mahacek Hunter, SCU

1. https://www.researchgate.net/figure/a-Martensite-and-austenite-crystal-structures-for-NiTi-SMA-and-b-pseudoelastic_fig5_287799881

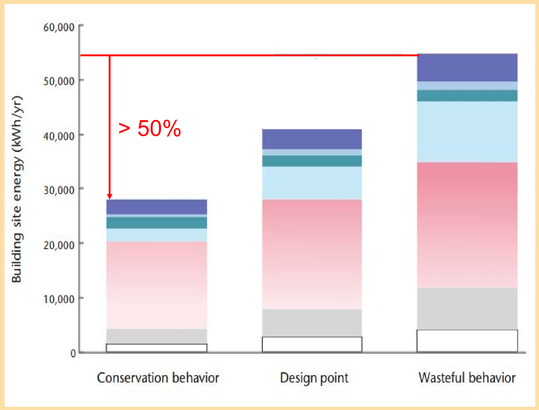
Multi-layer Sensor Data Processing for Occupancy Detection in Residential Buildings

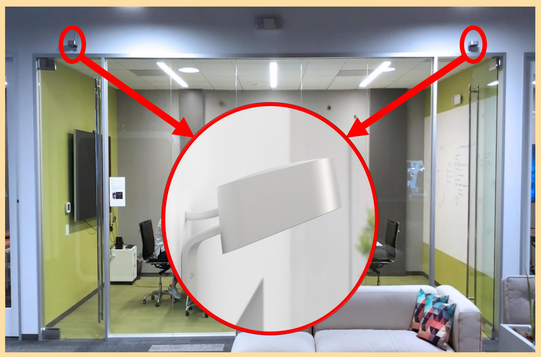
Chenli Wang (cwang@scu.edu), Hohyun Lee (hlee@scu.edu)
Mechanical Engineering Santa Clara University

Background



Energy conscious behaviors in residential homes reduces energy consumption by 33% compared to the design point, and half of energy consumption of those with wasteful behaviors. Examples of wasteful behaviors include overheating/cooling, ignoring malfunctioning equipment, or operating without occupants.





Existing Solution

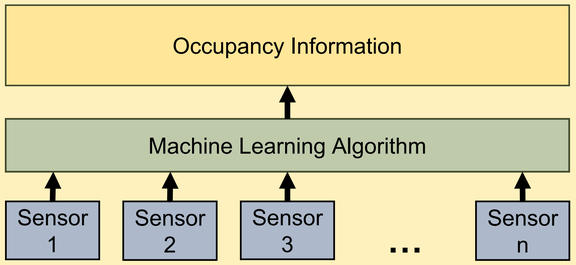
This table lists pros and cons of some existing occupancy measurement solutions on the market for residential buildings.

It can be seen that almost all of the individual detection systems are unable to provide quality occupancy information due to multiple reasons. Hence, we propose a new occupancy measurement system which is low-cost, accurate, easy to install and user-friendly.

	PIR	Electro -magnetic	CO_2	Camera
Presence	✓	✓	✓	✓
Location	✗	✗	✗	✓
Identity	✗	✓	✗	✓
Price	$20~$65	$40~$80	$80~$250	$50~$250
Disadvantages	Low Accuracy	User unfriendly	Expensive	Expensive Privacy

Machine Learning Based System

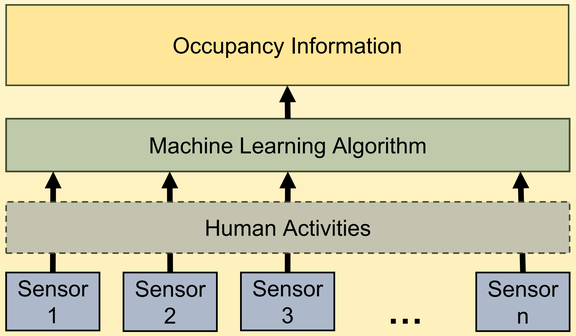
Machine learning (ML) approaches have been identified as an effective way to detect occupancy information as they can successfully extract significant information from the environment data as well as handle the randomness of human behaviors.



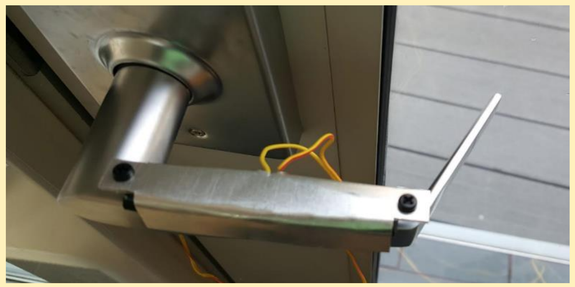
But the accuracy of ML models is highly dependent on the installation condition of the sensors, which is impossible to keep consistent in different residential buildings, posing great challenges to their general applications. Another challenge of ML is that it relies on sufficient data to extract appropriate features and train the model, which is time consuming and may increase the cost of the system.

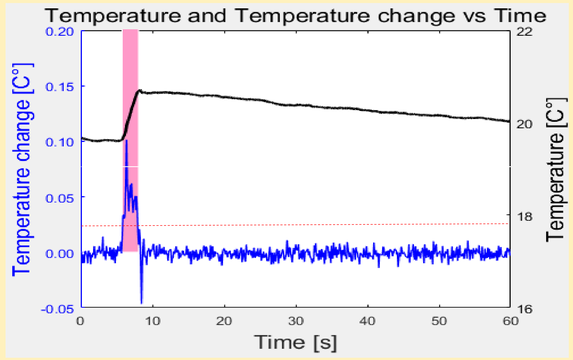
Incorporating Domain Knowledge

Incorporating domain knowledge has been proven an effective approach to help supervised and unsupervised machine learning models. In this work, instead of directly using sensor data to detect occupancy information, we use the data with event detection systems to retrieve everyday activities based on our domain knowledge (e.g., opening/closing door, using the toilet and using tap).



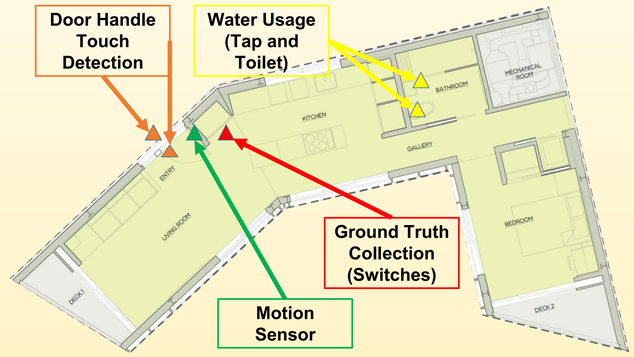
Touch Detection:





	Actual	Algorithm	Accuracy
Touch Test	300 Touches	296	98.7%
No touch test	Left on for two weeks	0	100%

Experiment Set-up



Results

Data: 10/18/2019 – 12/11/2019
Events: 489
80% Training , 20% Testing

	Standard ML		With Domain Knowledge	
	Decision Tree	Random Forest	Decision Tree	Random Forest
Accuracy	0.86	0.85	0.96	0.97
Precision	0.69	0.74	0.89	0.94
Recall	0.03	0.03	0.8	0.8
F1-score	0.06	0.06	0.84	0.86

Conclusion

This paper presents an easy-install, non-intrusive, economical, and multi-layer occupancy detection system in residential buildings. There are two different layers in the system: (1) a low layer event detection that retrieves occupants' activities (opening/closing door, using the toilet and using tap) from the environment data, and (2) an upper layer occupancy detection that utilizes machine learning models (Random Forest and Decision Tree) to determine the real-time occupancy information from different events.

The key contributions of our work are: (A) providing a tangible solution for occupancy measurement in residential buildings with limited number of low-cost and non-intrusive sensors, and (B) proposing a two-layer occupancy detection scheme with improved generality by incorporating domain knowledge into ML algorithms.

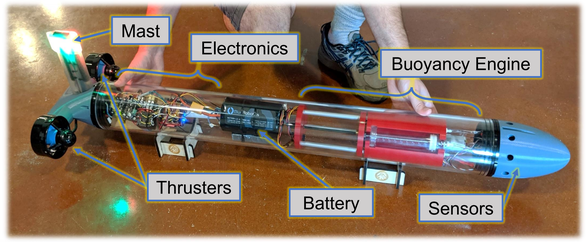
Development of a diving autonomous surface vessel and its applications in discrete sampling

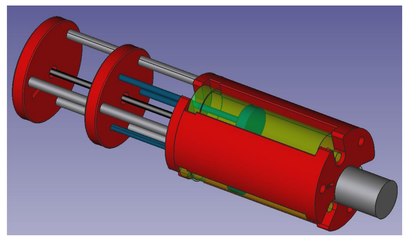
Max Woolsey and Dr. Christopher Kitts

Robotic Systems Lab, Department of Mechanical Engineering, Santa Clara University

Background

Undersea measurements are needed for monitoring events such as oil spills and low-oxygen "dead zones" which may not always have expressions of conditions at the surface. An economical marine vehicle was designed for implementing multi-robot adaptive sampling techniques in the field. This poster introduces the vehicle and sampling considerations that are currently being investigated.

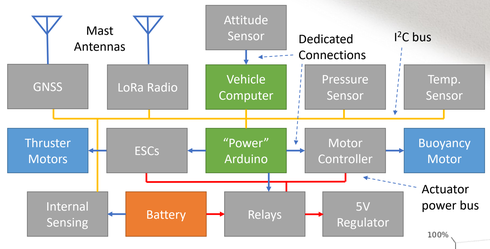




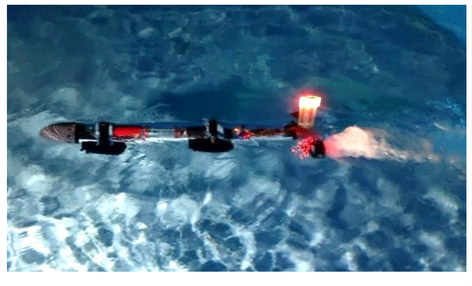
Variable Ballast system, "buoyancy engine" that intakes and expels 300mL of water

DASV

Diving Autonomous Surface Vehicle



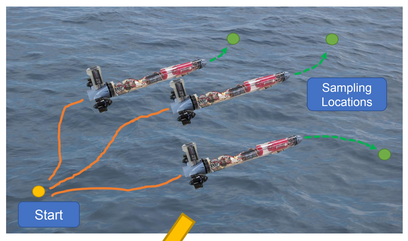
The diving autonomous surface vehicle (DASV) features a pair of thrusters for differential-drive control at the surface and a variable buoyancy system composed of actuated syringes allowing the vehicle to dive and surface for collecting profile measurements. At present, two prototype vehicles have been produced to develop different aspects of the overall DASV system.



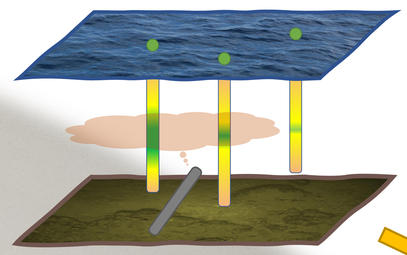
In-water actuator tests: transit (left), starboard turn (right)

Concept

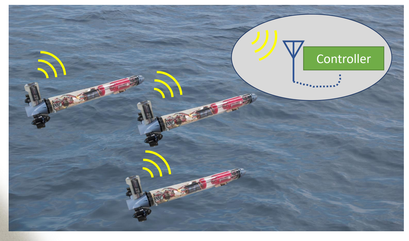
DASV discrete sampling concept of operation

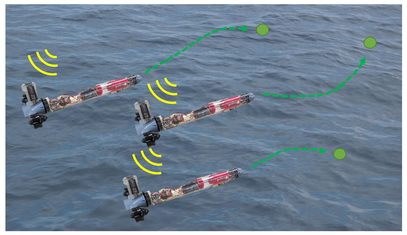


Multiple DASVs can be used to localize a subsurface feature. Each measurement cycle results in a set of vertical profiles, gradually building information about the feature.



An adaptive discrete sampling routine guides the vehicles to their sampling locations, and they dive to measure water column profiles. In this example, the profiles record contamination from a leaking pipeline.



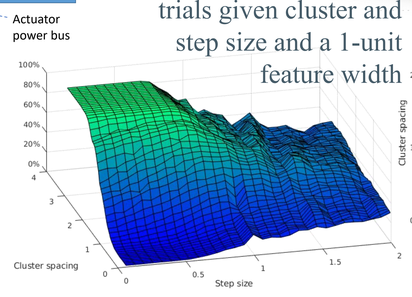


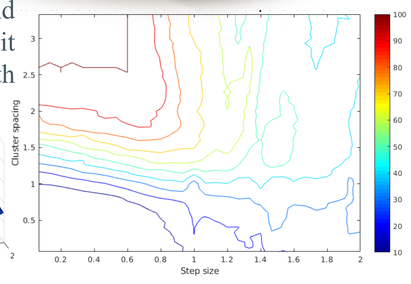
Observations are telemetered to the controller, which determines the next sampling points.

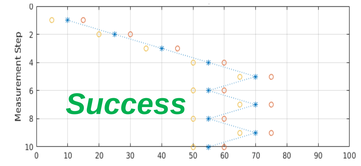
Simulation

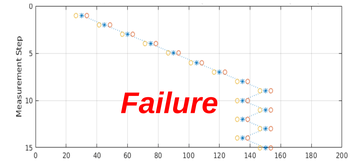
Adaptive Discrete Sampling in 1D

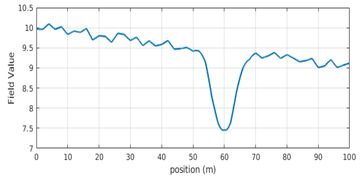
Percentage of successful trials given cluster and step size and a 1-unit feature width

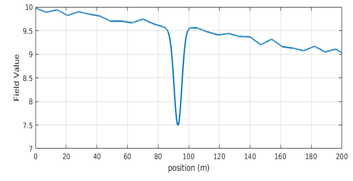












Simulations in 1D introduce the relationship between vehicle spacing (cluster size) and the sampling interval (step size), both relative to feature width. At each interval of the simulation, the pair of vehicles moves one step and measures the field; then the gradient is approximated to determine the direction of the next step. Field parameters are randomized, and each trial has three possible outcomes: the feature is located, the cluster stalls in a local minimum before reaching the feature, or the cluster passes the feature without detecting it. For this scenario, cluster size should be sufficiently wide to suppress detection of local minima, and step size should be less than ~75% of the feature width. This simulation does not account for the effort of sampling or transit; future routines may optimize a cost function for these items.

Modeling Bacterial Swimming in Biological Fluids

Noah Lordi
Engineering Physics
Santa Clara University

Ebru Demir
Beijing Computational Science
Research Center, Santa Clara University

Yang Ding
Beijing Computational Science
Research Science

On Shun Pak
Mechanical Engineering
Santa Clara University

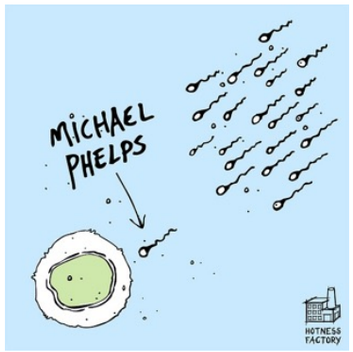
Introduction: Life at Low Reynolds Number

$$Re = \frac{\rho U L}{\mu} = \frac{\text{Inertial force}}{\text{Viscous force}}$$

$Re \sim 10^6$

$L \sim 1$ m

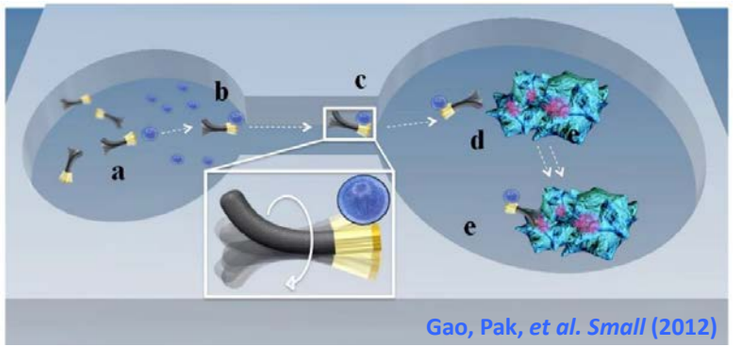
$Re \sim 10^{-4}$



$L \sim 1\ \mu$m

Biomedical Applications of Micro-swimmers

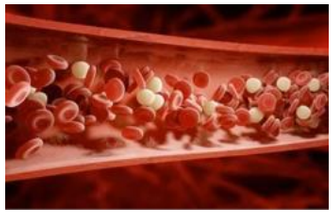
E.g., Targeted Drug Delivery, Micro-surgeries

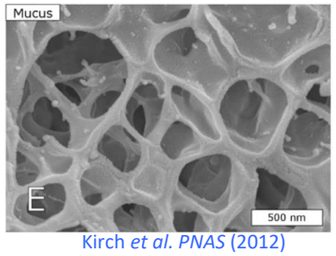


Gao, Pak, *et al. Small* (2012)

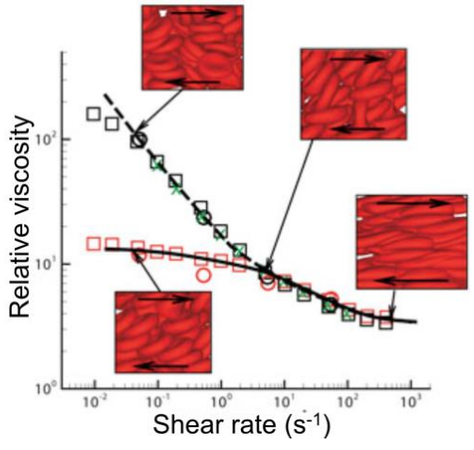
Locomotion in Non-Newtonian Fluids

- Shear-thinning viscosity is a ubiquitous non-Newtonian effect in biological fluids due to suspended microstructures.





Kirch *et al. PNAS* (2012)



Fedosov *et al. Biomech. Model. Mech.* (2013)

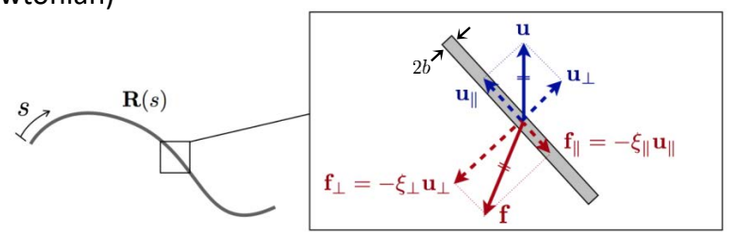
Goal: Impact of shear-thinning rheology on micro-propulsion

Local
(local reduction in viscosity)

- Resistive Force Theory (RFT)

RFT (Newtonian)

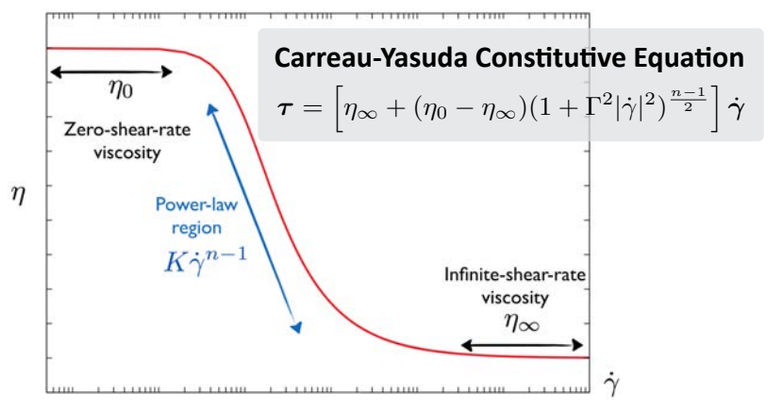
$$\mathbf{f} = -(\xi_\parallel \mathbf{u}_\parallel + \xi_\perp \mathbf{u}_\perp)$$



RFT (Shear-thinning)

$$\mathbf{f} = -R_C(\xi_\parallel \mathbf{u}_\parallel + \xi_\perp \mathbf{u}_\perp)$$

where $R_C = (1 + \Gamma^2 |\dot{\gamma}|^2)^{\frac{n-1}{2}}$

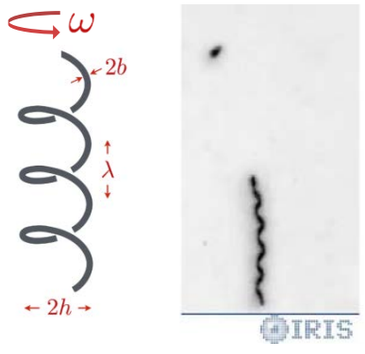


Carreau-Yasuda Constitutive Equation

$$\boldsymbol{\tau} = \left[\eta_\infty + (\eta_0 - \eta_\infty)(1 + \Gamma^2 |\dot{\gamma}|^2)^{\frac{n-1}{2}}\right] \dot{\boldsymbol{\gamma}}$$

Helical swimmer:

$$\mathbf{u} = \mathbf{u}_{\text{Strokes}} + \mathbf{U}_{\text{Swim}}$$
$$= h\omega\, \hat{\mathbf{e}}_\theta + U \hat{\mathbf{e}}_z$$



Uniform shear rate along the helix

$$\int_0^L R_C(\xi_\parallel \mathbf{u}_\parallel + \xi_\perp \mathbf{u}_\perp)\, \mathrm{ds} = 0$$

Nonlinear dependence on swimming speed

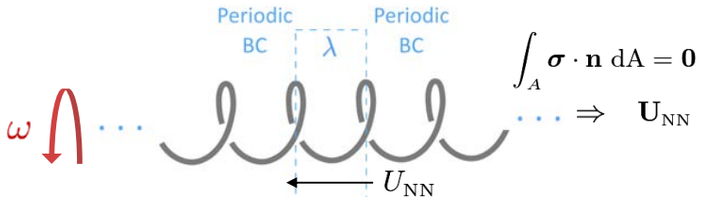
$$\Rightarrow R_C \int_0^L (\xi_\parallel \mathbf{u}_\parallel + \xi_\perp \mathbf{u}_\perp)\, \mathrm{ds} = 0$$

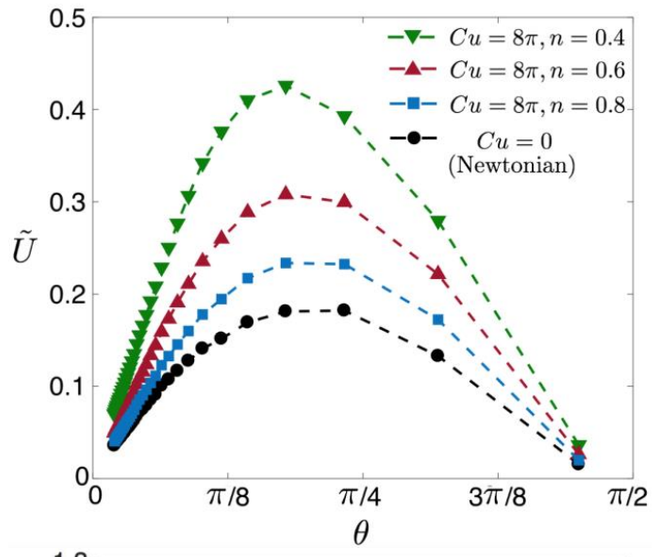
$$\Rightarrow \quad U_{\text{NN}} = U_{\text{N}}$$

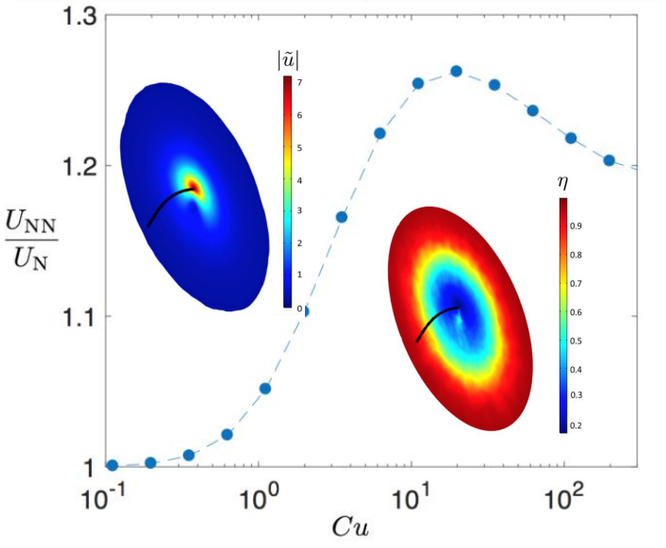
- **Local shear-thinning effect does not alter propulsion speed**

Non-Local
(induced flow change around the body)

Momentum equation: $\nabla p = \nabla \cdot \boldsymbol{\tau}$
Continuity equation $\nabla \cdot \mathbf{u} = 0$







- **Enhances propulsion speed significantly**
- **Reduces optimal helical angle**
- **Optimal Carreau number for propulsion**

Funding Acknowledgements: NSF CBET-1931292