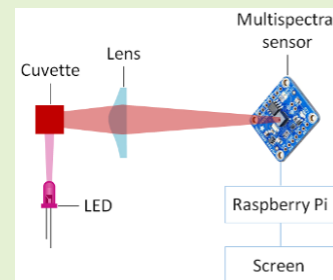


Multiplex Detection with Multispectral Sensors and Machine Learning

First A. Author, *Fellow, IEEE*, Second B. Author, and Third C. Author, Jr., *Member, IEEE*

Abstract—Multiplex assays usually use probes with different spectroscopic signatures to detect two or more analytes simultaneously. By allowing researchers to test for multiple markers per sample, they give a more reliable picture of what is happening. Being available in multiple colors and able to be conjugated to biomolecules, quantum dots are ideal as labels in multiplex assays. As a proof-of-concept, we demonstrate the capability of these multispectral sensors, using various machine learning approaches, for multiplex detection by using mixtures of quantum dots of two different colors. We show that the sensors can detect quantum dot concentrations relevant to biological applications by combining our sensor readings with ML algorithm running a Raspberry Pi Zero. We implement a 5 distinct types of ML algorithms to predict the QD concentration, along with various tools for training, evaluation, and utilizing a model.

Index Terms—Enter key words or phrases in alphabetical order, separated by commas. For a list of suggested keywords, send a blank e-mail to keywords@ieee.org or visit http://www.ieee.org/organizations/pubs/ani_prod/keywrdr98.txt



I. INTRODUCTION

COLORIMETRIC and luminescent assays are tests that generate colored reactants or fluorescence when an analyte of interest is present in a sample. They are widely used in biosensing, such as in disease diagnosis, food safety testing, environmental monitoring, and quality control. Since their signals are obtained through the spectral analysis of light, the detectors needed for these assays can be complex, bulky, and expensive. They often require trained personnel to maintain and operate them. These factors can slow, or even prevent, diagnosis in resource-limited regions. [1] Imaging-based technologies such as smartphone spectrometers have been touted as solutions to make biosensing accessible worldwide [2]. However, they need complicated procedures for calibrating the wavelength scale and converting digital images into spectra. (details about what is complicated about the procedures) Moreover, users have limited control over camera parameters. It restricts their

application to areas where people have the technical know-how to set them up properly in order to get reproducible image quality. Inexpensive multispectral sensors have recently been introduced commercially, and they have none of these constraints of imaging-based technologies. They require no calibration by the user as they come fitted with Gaussian interference filters and can monitor the following wavelengths simultaneously, with a full-width at half-maximum of 40 nm: 450 nm, 500 nm, 550 nm, 570 nm, 600 nm, and 650 nm. This set of wavelengths is adequate for many colorimetric, fluorescent, and luminescent assays. Being easy to assemble, troubleshoot and operate, these multispectral sensors are appealing to resource-limited laboratories. They have, for example, been interfaced with a Raspberry Pi Zero to produce a simple colorimetric reader for bio-assays. They are also versatile and can be found in applications ranging from wireless capsule endoscopy prototypes to systems characterizing microfluidic chips. In this work, we demonstrate their potential for multiplex detection. [1] kkk

This paragraph of the first footnote will contain the date on which you submitted your paper for review. It will also contain support information, including sponsor and financial support acknowledgment. For example, "This work was supported in part by the U.S. Department of Commerce under Grant BS123456."

The next few paragraphs should contain the authors' current affiliations, including current address and e-mail. For example, F. A. Author is with the National Institute of Standards and Technology, Boulder, CO 80305 USA (e-mail: author@boulder.nist.gov).

S. B. Author, Jr., was with Rice University, Houston, TX 77005 USA. He is now with the Department of Physics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: author@lamar.colostate.edu).

T. C. Author is with the Electrical Engineering Department, University of Colorado, Boulder, CO 80309 USA, on leave from the National Research Institute for Metals, Tsukuba, Japan (e-mail: author@nrim.go.jp).

A. Multiplex Applications of QDs

Multiplex assays usually use probes with different spectroscopic signatures to detect two or more analytes simultaneously. By allowing researchers to test for multiple markers per sample, they give a more reliable picture of what is happening. Consider, for example, a patient coming to a clinic with respiratory symptoms. The physician may take a sample from the patient, and if they have access to single-plex tests only, they will test for the pathogen they suspect is most likely responsible for the respiratory problem. If the test comes out

negative, they will need another sample and will have to test for another pathogen. In essence, the physician is using an elimination process for diagnosis. A multiplex test that screens the patient for multiple pathogens simultaneously will expedite the diagnosis process. Being available in multiple colors and able to be conjugated to biomolecules, quantum dots are ideal as labels in multiplex assays. They have, for example, been used to quantify the amount of cholera toxin, ricin, shiga-like toxin 1, and staphylococcal enterotoxin B in a sample. Multiplex analysis based on quantum dots have also been developed to identify illicit substances such as dexamethasone, gentamicin, clonazepam, medroxyprogesterone acetate, and ceftiofur. More recently, they have been applied to detect multiple tumor markers, or to monitor the activity of protease enzymes such as trypsin, chymotrypsin and enterokinase. As a proof-of-concept, we demonstrate the capability of these multispectral sensors, using various machine learning approaches, for multiplex detection by using mixtures of quantum dots of two different colors. We show that the sensors can detect quantum dot concentrations relevant to biological applications.

B. Machine learning and Sensors

In order to make predictions from a sensor input, we train statistical models on a training data set and test on unseen data to validate predictive power. We try various popular machine learning algorithms with different assumptions and requirements. Machine learning has been used for detection, quantification (regression task cite) and classification of past/current events [3], prediction of future ones or even self-calibration (Shirmohammadli2021 7&8).

In sensing applications, the ML algorithms are usually implemented digitally on general-purpose microprocessors [3], or (smart phone application cite). Both approaches have advantages and disadvantages. The notable limitations of a smart phone implementation is the difficulty redesigning applications and implementing new features. In this study we use Raspberry Pi (ARM Cortex A Series microprocessor), this is a cheap and widely accessible Single Board Computer (SBC). This device can alleviate some of the limitations of smart phone detection, and run a simple python programs with minimal set up. We implement 5 distinct types of ML algorithms to predict the QD concentration, along with various tools for training, evaluation, and utilizing a mode.

II. METHODS

A. Sensor System

Our system uses a commercially available multispectral sensor (Adafruit AS7341). Its data are read and processed with a standard I2C bus on an appropriate microcontroller, which was a Raspberry Pi Zero in our case. The system can be tailored to different applications since the software to run it is written in the Python language and can easily be modified. The microcontroller gives users the ability to attach a screen to view the data, or to transfer the results to another location via WiFi.

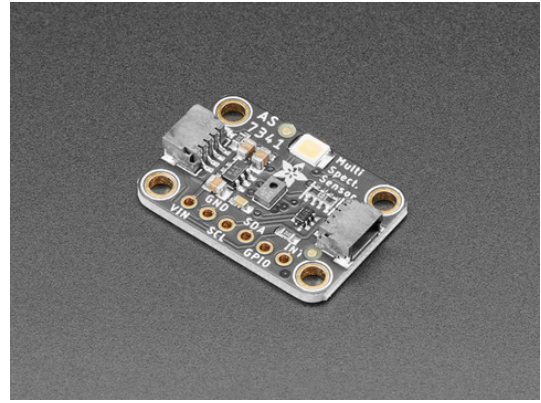


Fig. 1. Magnetization as a function of applied field. It is good practice to explain the significance of the figure in the caption.

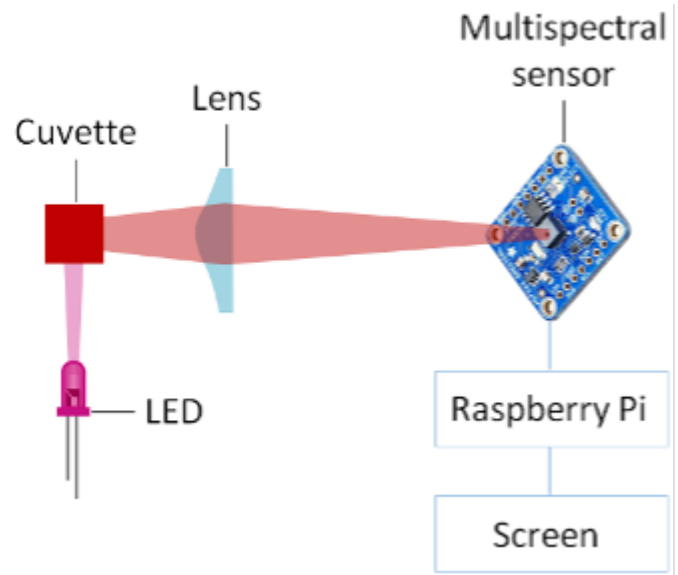


Fig. 2. Magnetization as a function of applied field. It is good practice to explain the significance of the figure in the caption.

B. Set up for testing our sensing system

The SI unit for magnetic field strength

C. Description of Machine learning algorithms

We use supervised machine learning to predict the QD concentration from the eight wavelength intensities. For each observation x_i for $i = 1, 2, \dots, n$ there is an associated response measurement y_i . The goal during training, is to approximate a mapping function from the input x to the output y variable, so that when given a new set of unknown input data, the algorithm can correctly predict the correct output variable. To avoid potential overfitting of the data, we have use a 10 fold-cross validation scheme. (some info about regression vs classification?) We have tried 5 distinct types of common ML algorithms to predict the QD concentration. In this section we provide a shot description and key properties of each ML algorithm. (following the style of Subhendu Pandit paper)

One possible approach to predict multiple QD types is to train a model for each emitter of interest, this approach is

flexible but more computationally and memory expensive. If using this scheme, during inference time we run all models in sequence. Alternatively, we could train our models to predict in a multivariate fashion, we discuss this possibility according to each algorithm.

Gaussian process regression: is nonparametric, meaning the data is not assumed to belong to any particular parametric family of probability distributions (cite x), so rather than calculating the probability distribution of parameters of a specific function, gaussian process regression calculates the probability distribution over all admissible functions that fit the data. However, we specify a function space prior, calculate the posterior using the training data, and compute the predictive posterior distribution on our points of interest. GP has been proven to be an effective method for nonlinear problems due to many desirable properties, such as a clear structure with Bayesian interpretation, a simple integrated approach of obtaining and expressing uncertainty in predictions and the capability of capturing a wide variety of data feature by hyperparameters [4]. The classical GP model can be only used to deal with a single output or single response problem however there are various ways to apply gaussian process regression with multiple response variables [4]

Regression MLP: Multi-level Perceptron (MLP) is a neural network connecting several layers of input nodes as a directed graph, tasks with predicting an output variable as a function of the inputs. The input features can be categorical or numeric types, however, for regression NN, we require a numeric dependent variable. The multiple layers and non-linear activation allow the MLP to learn non linear features. To make multivariate predictions we simply increase number of neurons in our output layer.

Random Forest: A random forest fits a number of classifying decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. Random forests also require less hyper-parameter tuning than other methods discussed here, with the exception of Linear regression. To make multivariate predictions we simply increase the output layer

SVM Regression: Support Vector Machine is a supervised learning algorithm that is used to predict discrete values. Support Vector Regression uses the same principle as the SVMs. A support-vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks like outliers detection (cite). Im actually not sure how to do multivariate for this example yet.

Linear regression: In a linear model, e.g. a model that assumes a linear relationship between the input variables (x) and the single output variable (y). More specifically, that y can be calculated from a linear combination of the input variables (x). When there is a single input variable, the method is referred to as simple linear regression otherwise we have multivariate example. Furthermore, to make multiple output predictions we extend linear regression however this can be shown to be same as independently solving a separate regression problem for each response variable.

D. Frequentist vs Bayesian Models

There are a few key differences in assumptions and applications between bayesian methods and frequentist methods. Both frequentist and Bayesian are statistical approaches to learning from data. However, Bayesian statistics concerns itself with trying to represent beliefs, the confidence intervals describing your initial (prior) belief and the update in your belief as you observe new data know as a posterior belief. The frequentist learning only depends on the observed data, frequentist statistics concerns itself with 'long run' (frequency) guarantees. For example, if we perform a sensor reading experiment on new data and create a 95% confidence interval around some quantity of interest then 95% of the readings will be in that confident interval, regardless of any prior expectations. In comparison, Bayesian method says the parameter's value is fixed but has been chosen from some probability distribution. we collect some data, and then calculate the probability of different values of the parameter given the data. This new probability distribution is called the "a posteriori probability" or simply the "posterior." Bayesian approaches can summarize their uncertainty by giving a range of values on the posterior probability distribution that includes 95% of the probability – this is called a 95% credibility interval.

E. Training Details

Here we can discuss general training details.

1) *Hyper-parameter Optimization:* Machine learning models have several hyperparameters that affect the corresponding optimization. "Often the general effects of hyperparameters on a model are known, but how to best set a hyperparameter and combinations of interacting hyperparameters for a given dataset is challenging." These can have a significant difference in learning, so its important we employ a search to find optimal hyper-parameters. In order to automate this process we create a GridSearch procedure for each one of the models attempted. Refer to the appendix for optimal hyperparameters found during training

2) *Confidence Intervals for retraining:* As discussed in section 3B. One of the major distinctions with Bayesian framework is is the probabilistic interpretation of a credibility interval. This allows for point-wise specific uncertainty, that can be used for flagging areas in the QD concentration range that have a high posterior uncertainty.

III. RESULTS

A. Characterizing the limit of detection

A conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

B. Model Selection

1) *Metrics:* A Cost Function is used to evaluate the performance of a Model on data. Mean squared error (MSE) is a common measure for evaluating the performance of a model which sums (or averages) the square of the difference between

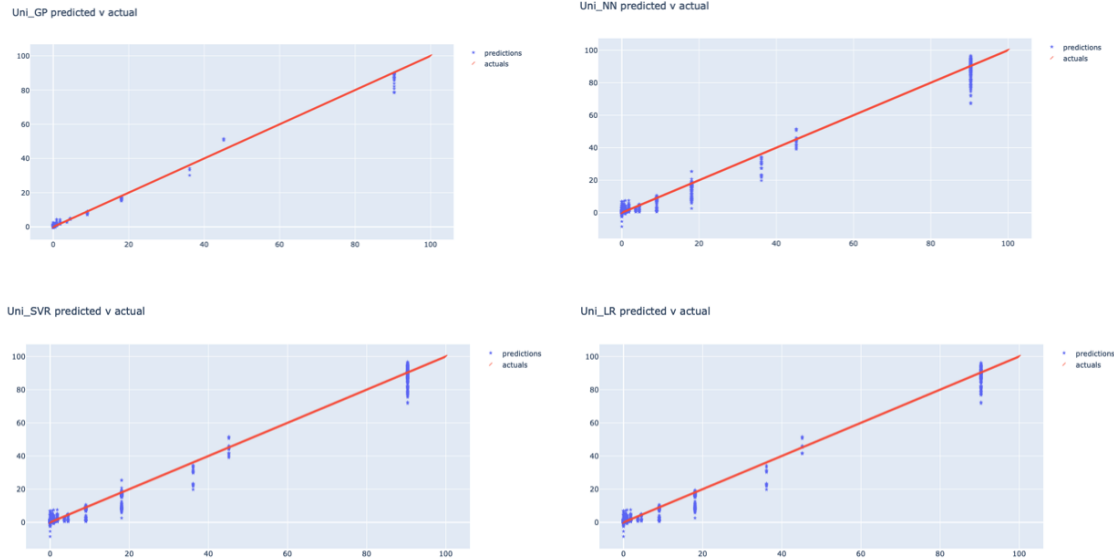


Fig. 3. Magnetization as a function of applied field. It is good practice to explain the significance of the figure in the caption.

TABLE I
MACHINE LEARNING MODELS

Model	Type	Accuracy*	Trainable Parameters
Gaussian Process Regression	Univariate	8.5	N/A**
Linear Regression	Univariate	4.5	2
Bayesian Regression	Univariate	3.5	N/A**
Random Forest Regression	Univariate	1.1	2400
Neural Network	Univariate	9.0	340
Neural Network	Multiivariate	15.0	2400

*Accuracy is measure of MSE for 10-fold Cross val. All models were tested using the same cross validation dataset breakdown. For univariate models accuracy is the average between the two models

**Bayesian methods are non-parametric and therefore have no trainable parameters

estimated and true values over the population. The main motivation for using MSE is its differentiability which simplifies the Cost Function minimization process. An alternative metric is the Mean absolute error (MAE) (i.e., the sum/average of absolute values of the differences between estimations and observed outputs) which is sometimes employed for sequential/temporal data processing. As mentioned before, training is achieved through by adjusting the weights (i.e., elements of vector \mathbf{w} in Eq. 1) \mathbf{w} .

Here we discuss cross validation accuracy and BIC, AIC criteria?

2) *Reporting Uncertainty*: Figures that are composed of only black lines and shapes. These figures should have no shades or half-tones of gray, only black and white.

C. Making Predictions

Raspberry Pi Details: After training a model, we do a running mean of the last 5 sensor readings, and run inference on the selected model of choice. It is important to fix the the placement and calibration of the sensor according to the

settings during data collection. We store the raw data and predictions in a csv file.

IV. CONCLUSION

Multiplex assays usually use probes with different spectroscopic signatures to detect two or more analytes simultaneously. By allowing researchers to test for multiple markers per sample, they give a more reliable picture of what is happening. Being available in multiple colors and able to be conjugated to biomolecules, quantum dots are ideal as labels in multiplex assays. As a proof-of-concept, we demonstrate the capability of these multispectral sensors, using various machine learning approaches, for multiplex detection by using mixtures of quantum dots of two different colors. We show that the sensors can detect quantum dot concentrations relevant to biological applications by combining our sensor readings with ML algorithm running a Raspberry Pi Zero. We implement a 5 distinct types of ML algorithms to predict the QD concentration, along with various tools for training, evaluation, and utilizing a model.

Appendixes, if needed, appear before the acknowledgment.

ACKNOWLEDGMENT

The preferred spelling of the word “acknowledgment” in American English is without an “e” after the “g.” Use the singular heading even if you have many acknowledgments. Avoid expressions such as “One of us (S.B.A.) would like to thank” Instead, write “F. A. Author thanks” In most cases, sponsor and financial support acknowledgments are placed in the unnumbered footnote on the first page, not here.

REFERENCES AND FOOTNOTES

A. References

Other than books, capitalize only the first word in a paper title, except for proper nouns and element symbols. For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation. See the end of this document for formats and examples of common references. For a complete discussion of references and their formats, see the IEEE style manual at <http://www.ieee.org/authortools>.

APPENDIX I

SUBMITTING YOUR PAPER FOR REVIEW

A. Final Stage

When you submit your final version (after your paper has been accepted), print it in two-column format, including figures and tables. You must also send your final manuscript on a disk, via e-mail, or through a Web manuscript submission system as directed by the society contact. You may use *Zip* for large files, or compress files using *Compress*, *Pkzip*, *Stuffit*, or *Gzip*.

APPENDIX II

PUBLICATION PRINCIPLES

The two types of contents of that are published are; 1) peer-reviewed and 2) archival. The Transactions and Journals Department publishes scholarly articles of archival value as well as tutorial expositions and critical reviews of classical subjects and topics of current interest.

Authors should consider the following points:

- 1) Technical papers submitted for publication must advance the state of knowledge and must cite relevant prior work.
- 2) The length of a submitted paper should be commensurate with the importance, or appropriate to the complexity, of the work. For example, an obvious extension of previously published work might not be appropriate for publication or might be adequately treated in just a few pages.
- 3) Authors must convince both peer reviewers and the editors of the scientific and technical merit of a paper; the standards of proof are higher when extraordinary or unexpected results are reported.
- 4) Because replication is required for scientific progress, papers submitted for publication must provide sufficient information to allow readers to perform similar experiments or calculations and use the reported results. Although not everything need be disclosed, a paper must contain new, useable, and fully described information. For example, a specimen's chemical composition need not be reported if the main purpose of a paper is to introduce a new measurement technique. Authors should expect to be challenged by reviewers if the results are not supported by adequate data and critical details.
- 5) Papers that describe ongoing work or announce the latest technical achievement, which are suitable for presentation at a professional conference, may not be appropriate for publication.

APPENDIX III

REFERENCE EXAMPLES

REFERENCES

- [1] D. Wang, F. Lian, S. Yao, Y. Liu, J. Wang, X. Song, L. Ge, Y. Wang, Y. Zhao, J. Zhang, C. Zhao, and K. Xu, "Simultaneous detection of three foodborne pathogens based on immunomagnetic nanoparticles and fluorescent quantum dots," *ACS Omega*, vol. 5, no. 36, pp. 23070–23080, 2020.
- [2] F. A. Esteve-Turrillas and A. Abad-Fuentes, "Applications of quantum dots as probes in immunosensing of small-sized analytes," *Biosensors and Bioelectronics*, vol. 41, no. 1, pp. 12–29, 2013.
- [3] V. Shirmohammadi and B. Bahreyni, "Machine Learning for Sensing Applications: A Tutorial," *IEEE Sensors Journal*, vol. XX, no. XX, pp. 1–1, 2021.
- [4] Y. Zhao, M. Ye, Q. Chao, N. Jia, Y. Ge, and H. Shen, "Simultaneous detection of multiood-borne pathogenic bacteria based on functionalized quantum dots coupled with immunomagnetic separation in food samples," *Journal of Agricultural and Food Chemistry*, vol. 57, no. 2, pp. 517–524, 2009.



First A. Author (M'76–SM'81–F'87) and all authors may include biographies. Biographies are often not included in conference-related papers. This author became a Member (M) of IEEE in 1976, a Senior Member (SM) in 1981, and a Fellow (F) in 1987. The first paragraph may contain a place and/or date of birth (list place, then date). Next, the author's educational background is listed. The degrees should be listed with type of degree in what field, which institution, city, state, and country, and year the degree was

earned. The author's major field of study should be lower-cased.

The second paragraph uses the pronoun of the person (he or she) and not the author's last name. It lists military and work experience, including summer and fellowship jobs. Job titles are capitalized. The current job must have a location; previous positions may be listed without one. Information concerning previous publications may be included. Try not to list more than three books or published articles. The format for listing publishers of a book within the biography is: title of book (publisher name, year) similar to a reference. Current and previous research interests end the paragraph. The third paragraph begins with the author's title and last name (e.g., Dr. Smith, Prof. Jones, Mr. Kajor, Ms. Hunter). List any memberships in professional societies other than the IEEE. Finally, list any awards and work for IEEE committees and publications. If a photograph is provided, it should be of good quality, and professional-looking. Following are two examples of an author's biography.



Second B. Author was born in Greenwich Village, New York, NY, USA in 1977. He received the B.S. and M.S. degrees in aerospace engineering from the University of Virginia, Charlottesville, in 2001 and the Ph.D. degree in mechanical engineering from Drexel University, Philadelphia, PA, in 2008.

From 2001 to 2004, he was a Research Assistant with the Princeton Plasma Physics Laboratory. Since 2009, he has been an Assistant Professor with the Mechanical Engineering Department, Texas A&M University, College Station. He is the author of three books, more than 150 articles, and more than 70 inventions. His research interests include high-pressure and high-density nonthermal plasma discharge processes and applications, microscale plasma discharges, discharges in liquids, spectroscopic diagnostics, plasma propulsion, and innovation plasma applications. He is an Associate Editor of the journal *Earth, Moon, Planets*, and holds two patents.

Dr. Author was a recipient of the International Association of Geomagnetism and Aeronomy Young Scientist Award for Excellence in 2008, and the IEEE Electromagnetic Compatibility Society Best Symposium Paper Award in 2011.



Third C. Author, Jr. (M'87) received the B.S. degree in mechanical engineering from National Chung Cheng University, Chiayi, Taiwan, in 2004 and the M.S. degree in mechanical engineering from National Tsing Hua University, Hsinchu, Taiwan, in 2006. He is currently pursuing the Ph.D. degree in mechanical engineering at Texas A&M University, College Station, TX, USA.

From 2008 to 2009, he was a Research Assistant with the Institute of Physics, Academia Sinica, Taipei, Taiwan. His research interest includes the development of surface processing and biological/medical treatment techniques using nonthermal atmospheric pressure plasmas, fundamental study of plasma sources, and fabrication of micro- or nanostructured surfaces.

Mr. Author's awards and honors include the Frew Fellowship (Australian Academy of Science), the I. I. Rabi Prize (APS), the European Frequency and Time Forum Award, the Carl Zeiss Research Award, the William F. Meggers Award and the Adolph Lomb Medal (OSA).