

# Point-process based representation learning for Electronic Health Records

Hojjat Karami, *Fellow, IEEE*, Anisoara Ionescu, and David Atienza, *Member, IEEE*

**Abstract**—Irregular @hornSetFunctionsTime2020 sampling of time series in electronic health records (EHRs) is challenging for model development. In addition, the pattern of missingness for certain clinical variables are not at random as it is determined by clinicians decision and the state of patient. Point process is a mathematical framework for handling event sequence data which is also consistent with the irregular sampling. We propose *TEEDAM*, a deep neural network that can learn patient representation from irregularly sample time series as well as informative missingness pattern of certain laboratory variables. We performed various experiments to show the effectiveness of event and state encoding for characterization of conditional intensity functions as well as downstream prediction task. Results show that in some cases learning from patterns may improve the performance of prediction task.

**Index Terms**—electronic health records (EHRs), Point Process, irregular sampling, informative missingness

## I. INTRODUCTION

healthcare systems are one of main area for deploying machine learning models. Thanks to recent advancements in data collection technologies, policies and computing power data are recorded in a hospital. This data consists of multiple modalities from tabular and time series to clinical notes and images. We focus on tabular and time series.

talk about the irregular sampling in EHRs and causes./// In and ideal setup, variables are measured on a regular basis, however, in a hospital due to medical costs, patient associated risks, sampling device.

conseq of irr sampling///

potential adv of irr sampl///

imputation methods///

inherently compatible methods///

pp framework///

Our aims?/// In this work, we propose a DNN for joint modeling of NTPP and a DAM. Our first aim is to investigate the effectiveness of state encoding for CIF characterization.

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 by DigiPredict Grant BS123456."

H. Karami is with the EPFL, Lausanne, Switzerland (e-mail: hojjat.karami@epfl.ch).

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).

Second, we are interested in seeing the utility of event encoding a downstream task  
contributes?///

## II. BACKGROUND

1) *Temporal point process*: Consider an event sequence data  $\mathcal{D} = \{\mathcal{S}_i\}_{i=1}^N$  where each sample is represented as an event sequence  $\mathcal{S}_i = \{(t_i, e_i)\}_{j=1}^L$ , where  $L$  is number of event,  $t_j$  is event's timestamp, and  $e_j \in \mathbb{R}^M$  is the binary representation of event marks (multi-class or multi-label). The history of events is denoted as  $\mathcal{H}_t = \{(t_j, e_j) : t_j < t\}$

In the multi-class setting where each mark can be exactly one of  $M$  marks, the point process is characterized by  $M$  conditional intensity functions (CIFs):

$$\lambda_m^*(t) = \lim_{\Delta t} \frac{P(\text{event of type } m \text{ in } [t, t + \Delta t] | \mathcal{H}_t)}{\Delta t} \quad (1)$$

The default negative log-likelihood (NLL) of a point process for a single event sequence  $\mathcal{S}_i$  is defined as:

$$-\log p(\mathcal{S}_i) = -\sum_{j=1}^L \sum_{m=1}^M 1(e_j = m) \log \lambda_m^*(t_j) + \sum_{m=1}^M \left( \int_0^T \lambda_m^*(s) ds \right) \quad (2)$$

If we assume that marks and events are independent,

$$-\log p(\mathcal{S}_i) = -\sum_{j=1}^L \sum_{m=1}^M 1(e_j = m) \log p^*(e_j = m) \lambda_m^*(t_j) + \sum_{m=1}^M p^*(e_j = m) \left( \int_0^T \lambda_m^*(s) ds \right) \quad (3)$$

[ntpp] has proposed the following CIF to extend for multi-label case:

$$\begin{aligned}
-\log p(\mathcal{S}_i) = & - \sum_{j=1}^L \sum_{m=1}^M 1(e_j = m) \log \lambda_m^*(t_j) \\
& + \sum_{m=1}^M \left( \int_0^T \lambda_m^*(s) ds \right) \\
& - \sum_{j=1}^L \sum_{m=1}^M (1 - 1(e_j = m)) \log \left( 1 - \lambda_m^*(t_j) \exp \left( - \int_0^T \lambda_m^*(s) ds \right) \right)
\end{aligned}$$

It should be noted that the real advantage of point process is modeling non-event likelihoods in the form of integrals. If we neglect the integrals, we would achieve the cross-entropy and binary cross entropy loss in the multi-class and multi-label settings respectively for the prediction of next mark given history of events.

2) *Neural temporal point process*: Encoder-decoder architectures have proven to be effective in many applications. The main idea of a neural point process is to first encode the history of events until  $t_j$  using different neural network architectures  $h_j = \text{Enc}(\mathcal{H}_j)$ . Then it tries to estimate  $\lambda_m^*(t)$  or its integral for  $t_i < t \leq t_{i+1}$ .

3) *Deep learning for irregular sampling*: An irregularly sampled data can be denoted as  $\mathcal{D} = \{\mathcal{U}_i\}_{i=1}^N$  where  $N$  is number of samples and each sample is represented as series of tuple  $\mathcal{U}_i = (t_p, k_p, v_p)$  where  $t_p, k_p, v_p$  represents the time, modality and value of  $p$ -th datapoint respectively.

### III. RELATED WORKS

#### IV. PROPOSED MODEL

The key advantage of our proposed model is to combine a transformer-based event encoder (TEE) with point process loss with a deep attention module (DAM) that can handle an irregularly sampled time series for a any downstream prediction task. In this situation the data is represented as  $\mathcal{D} = \{\mathcal{S}_i, \mathcal{U}_i\}_{i=1}^N$ . General schematic of TEDAM is depicted in FIG.

#### A. Event Encoder

We use a similar transformer architecture [thp] for encoding events with minor modifications for event embedding.

In the first step, we embed all event marks  $E_{emb} = E \times W_{emb}$  where  $E_i \in \mathbb{R}^{L \times M}$  is the binary encoding matrix of all event marks (multi-label or multi-class), and  $W_{emb} \in \mathbb{R}^{M \times d_{emb}}$  is the trainable embedding matrix.

In the second step, timestamps should be encoded and added to the event embedding, however, we propose to concatenate time encodings that can lead to better characterization of conditional intensity functions. Finally, the input of the transformer encoder will be  $X_{emb} = [E_{emb}, T_{emb}] \in \mathbb{R}^{L \times (d_{emb} + t_{emb})}$ .

Here, we use the standard transformer encoder similar to [vaswani] with masking matrix to prevent the model from looking into the future. we obtain the encoded matrix  $H = (h_1, \dots, h_j, \dots, h_L)$  where  $h_j$  contains the all available information before occurrence of  $j$ -th event.

#### B. State Encoder

Similar to [setF], we use an attention-based aggregation approach for encoding all additional information. Each side information  $(t_k, v_k, m_k)$  can be represented by  $s_k = (z(t_k), v_k, m_k)$ . we define attention  $a(\mathcal{S}_k, s_k)$

We define  $\mathcal{S}_p$  to be the set of the first  $p$  available information. The goal is to calculate  $a(\mathcal{S}_p, s_k), k \leq p$  that is the relevance of  $k$ -th observation  $s_k$  to the first  $p$  observed values  $\mathcal{S}_p$ . This is achieved by computing an embedding of the set elements using a smaller set functions  $f'$ , and projecting the concatenation of the set representation and the individual set element into d-dimensional space:

The goal is to calculate  $a(\mathcal{S}_p, s_k), k \leq p$  that is the relevance of  $k$ -th observation  $s_k$  to the first  $p$  observed values  $\mathcal{S}_p$ .

This is achieved by embedding set elements  $\mathcal{S}_p$  using a simpler set function  $f'$  such as a cumulative mean.

$$f'(\mathcal{U}_p) = g' \left( \frac{1}{|p|} \sum_{u_k \in \mathcal{U}_p} h'(u_k; \theta'); \rho' \right) \quad (5)$$

Then we can compute key values using key matrix  $W^k \in \mathbb{R}^{(d_g + d_s) \times d_{prod}}$

$$K_p = [f'(\mathcal{U}_p), u_p]^T W^K \quad (6)$$

using a query vector  $w^q \in \mathbb{R}^{d_{prod}}$

$$a(\mathcal{U}_p, u_p) = \text{softmax} \left( \frac{K_p \cdot w^q}{\sqrt{d}} \right)$$

Finally, we compute a weighted aggregation of set elements:

$$z_p = f(\mathcal{U}_p) = g \left( \sum_{s_k \in \mathcal{U}_p} a(\mathcal{U}_p, s_k) h(s_k; \theta); \rho \right)$$

we regard  $z_p \in \mathbb{R}^{d_p}$  as the representation of state data until arrival of  $p$ -th datapoint.

Finally, we need to combine event embeddings  $H_{L \times d_e}$  and state embeddings  $Z_{P \times d_p}$ , however, the length of each does not match. As a result, we consider the reduced version of state matrix as below:

$$Z_i^{red} = Z_p \text{ where } p = \text{argmax}(z \leq i)$$

Without loss of generality, we can consider multiple heads by adding an additional dimension to keys and queries.

All formulas are:

$$\begin{cases}
f'(\mathcal{S}_p) = g' \left( \frac{1}{p} \sum_{s_k \in \mathcal{S}_p} h'_\theta(s_k) \right) \\
K_p = [f'(\mathcal{S}_p), s_p]^T W^K \\
e_p = \frac{K_p \cdot w^q}{\sqrt{d}} \\
a(\mathcal{S}_p, s_k) = \frac{\exp(e_p)}{\sum_{k \leq p} \exp(e_k)} \\
f(\mathcal{S}_p) = \sum_{s_k \in \mathcal{S}_p} a(\mathcal{S}_p, s_k) h_\theta(s_k) \\
z_p = g_\psi(f(\mathcal{S}_p))
\end{cases}$$

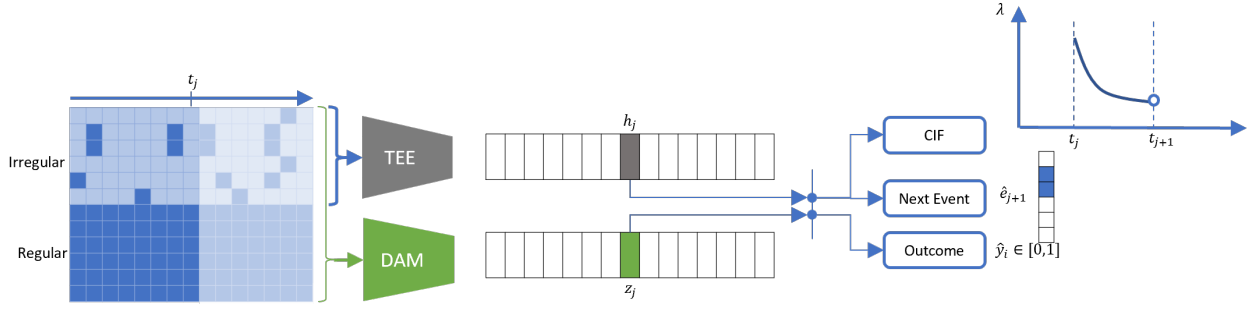


Fig. 1. Magnetization as a function of applied field. It is good practice to explain the significance of the figure in the caption.

$$[z(t_j)]_k = \begin{cases} \sin\left(\frac{t_j}{\mathcal{T}^{(k-1)}/d_t}\right) & \text{if } k \text{ is odd} \\ \sin\left(\frac{t_j}{\mathcal{T}^k/d_t}\right) & \text{if } k \text{ is even} \end{cases} \quad (7)$$

$$\lambda_k(t|\mathcal{H}_t) = f_k\left(\alpha_k \frac{t-t_j}{t_j} + \mathbf{w}_k^T \mathbf{x}_{enc}(t_j) + \mathbf{y}_k^T \mathbf{s}_{enc}(t_j) + b_k\right) \quad (8)$$

$$\lambda_k(t|\mathcal{H}_t) = f_k\left(\alpha_k \frac{t-t_j}{t_j} + \mathbf{w}_k^T \mathbf{x}_{enc}(t_j) + b_k\right) \quad (9)$$

### C. Event Decoder

Once we obtain a representation of a patient using embedded events and states, we can try to parameterize conditional intensity functions (CIFs) of the events.

In neural point process literature, many approaches have been propose to decode either conditional or cumulative intensity function. We will use a decoder similar to [sahp] as it can model both exciting and inhibiting effects for modeling CIFs.

$$\begin{aligned} \mu_{m,i+1} &= \text{gelu}(h_{i+1}W_{m,\mu} + z_{i+1}W_{m,\mu}), \\ \eta_{m,i+1} &= \text{gelu}(h_{i+1}W_{m,\eta} + z_{i+1}W_{m,\eta}), \\ \gamma_{m,i+1} &= \text{gelu}(h_{i+1}W_{m,\gamma} + z_{i+1}W_{m,\gamma}), \end{aligned}$$

Finally, we can express the intensity function as follows:

$$\lambda_m(t) = \text{softplus}(\mu_{m,i+1} + (\eta_{m,i+1} - \mu_{m,i+1}) \exp(-\gamma_{m,i+1}(t - t_i))),$$

for  $t \in (t_i, t_{i+1}]$ , where the *softplus* is used to constrain the intensity function to be positive.

### D. Loss Function

We define a multi-objective loss function  $\mathcal{L} = \mathcal{L}_{CIF} + \mathcal{L}_{mark} + \mathcal{L}_{state}$ .

## V. EXPERIMENTS

We perform various experiments to show the effectiveness of each component in our model.

### Datasets

To show the utility of time concatenation and marked loss, we consider three datasets:

**Synthea (Syn).** We used the Synthea simulator (Walonoski et al., 2018) which generates patient-level EHRs using human expert curated Markov processes. Here, we reused the already processed version of this data by [ntpp].

**ReTweets (RT).** The Retweets dataset contains sequences of tweets, where each sequence contains an origin tweet (i.e., some user initiates a tweet), and some follow-up tweets. We record the time and the user tag of each tweet. Further, users are grouped into three categories based on the number of their followers: “small”, “medium”, and “large”

**Stackoverflow (SO).** is a question-answering website. The website rewards users with badges to promote engagement in the community, and the same badge can be rewarded multiple times to the same user. We collect data in a two-year period, and we treat each user’s reward history as a sequence. Each event in the sequence signifies receipt of a particular medal.

Furthermore, we consider two EHRs provided by physionet challenge to investigate the advantage of irregular sample and point process modeling in the same time.

**Physionet 2012 Mortality Prediction Challenge (P19).** The 2012 Physionet challenge dataset (Goldberger et al., 2000), contains 12,000 ICU stays each of which lasts at least 48 h. For each stay, a set of general descriptors (such as gender or age) are collected at admission time. Depending on the course of the stay and patient status, up to 37 time series variables were measured (e.g. blood pressure, lactate, and respiration rate). While some modalities might be measured in regular time intervals (e.g. hourly or daily), some are only collected when required; moreover, not all variables are available for each stay.

**Physionet 2019 Sepsis Early Prediction Challenge (P19).** This dataset contains clinical data of about 40k patients in ICU. Clinical data consist of demographics, vital signs and laboratory values as well as sepsis label in a one-hour time grid. Our objective is to predict the timestamp of next lab sampling events as well as measured variables (event marks) given the patient history.

### Scenarios

To show the utility of time concatenation and marked-shp we report the metrics for SO, RT and SYN and compare it with 3 baselines: SAHP, THP, and GRU-CP.

To show the utility of additional information for CIF modeling, we report NLL/events for P12 and P19 in the following conditions: TE, TE+DAM, TE+noise

Finally, we investigate whether point process characterization can be useful in a downstream task or not. As each dataset consists of different hospitals and the pattern of irregular sampling might differ from hospital to hospital, we assume different settings: single-center, multi-center and external evaluation.

### Baselines

we compare our model against existing models: THP, SAHP, GRU-CP.

### Metrics

We report the weighted AUPRC, AUROC of next predicted event as well as root mean square error (RMSE) of next measurement interval. For evaluating the goodness of fit for the parameterized point process, we report normalized negative likelihood normalized by number of occurred event (NLL/events). Furthermore, we can also evaluate the learned representation of each patient to predict the sepsis label in a binary classification task.

### Training Details

To be completed.

## VI. RESULTS AND DISCUSSION

In this section, we present our results regarding the advantage of state and event encoding.

### A. Effect of minor improvements

effect of time concatenation  
compare single+mark with mc or ml

### B. Negative Likelihood with state encoding

Table 1 shows the result for estimation of negative likelihood in different datasets and scenarios. It is obvious that state encoding has led to lower NLL.

can u provide one example patient?

### C. Downstream task with event encoding

Another key element of our work is to show the effectiveness of point process modeling for a downstream task. In Table, we have reported the performance metrics for the mortality prediction task. We have compared our results with several sota's DL models that are compatible with irregular time series.

### D. Learned representations

Fig 1 visualizes the tsne plot for the two scenarios.

### E. Model interpretability

one advantage of proposed method is use of attention mechanisms in both event and state encoder. Fig 1 shows the attention mechanism

### F. Likelihood estimation

Although CIF does not improve mark prediction, it has led to better representation of patient for downstream task such as sepsis prediction.

In addition, we can interpret some of learned CIF patterns. explain the effect of time concatenation in SO dataset tsne of learned representation. 4 modes:

- (DA,TE)- $\hat{c}$ (Mark, CIF)
- attention of DA for sepsis prediction  
attention matrix of events for SO dataset

## VII. CONCLUSION

### REFERENCES

- [1] G. O. Young, "Synthetic structure of industrial plastics," in *Plastics*, 2<sup>nd</sup> ed., vol. 3, J. Peters, Ed. New York, NY, USA: McGraw-Hill, 1964, pp. 15–64.
- [2] W.-K. Chen, *Linear Networks and Systems*. Belmont, CA, USA: Wadsworth, 1993, pp. 123–135.
- [3] J. U. Duncombe, "Infrared navigation—Part I: An assessment of feasibility," *IEEE Trans. Electron Devices*, vol. ED-11, no. 1, pp. 34–39, Jan. 1959, 10.1109/TED.2016.2628402.
- [4] E. P. Wigner, "Theory of traveling-wave optical laser," *Phys. Rev.*, vol. 134, pp. A635–A646, Dec. 1965.
- [5] E. H. Miller, "A note on reflector arrays," *IEEE Trans. Antennas Propagat.*, to be published.
- [6] E. E. Reber, R. L. Michell, and C. J. Carter, "Oxygen absorption in the earth's atmosphere," Aerospace Corp., Los Angeles, CA, USA, Tech. Rep. TR-0200 (4230-46)-3, Nov. 1988.
- [7] J. H. Davis and J. R. Cogdell, "Calibration program for the 16-foot antenna," Elect. Eng. Res. Lab., Univ. Texas, Austin, TX, USA, Tech. Memo. NGL-006-69-3, Nov. 15, 1987.
- [8] *Transmission Systems for Communications*, 3<sup>rd</sup> ed., Western Electric Co., Winston-Salem, NC, USA, 1985, pp. 44–60.
- [9] *Motorola Semiconductor Data Manual*, Motorola Semiconductor Products Inc., Phoenix, AZ, USA, 1989.
- [10] G. O. Young, "Synthetic structure of industrial plastics," in *Plastics*, vol. 3, Polymers of Hexadromicon, J. Peters, Ed., 2<sup>nd</sup> ed. New York, NY, USA: McGraw-Hill, 1964, pp. 15–64. [Online]. Available: <http://www.bookref.com>.
- [11] *The Founders' Constitution*, Philip B. Kurland and Ralph Lerner, eds., Chicago, IL, USA: Univ. Chicago Press, 1987. [Online]. Available: <http://press-pubs.uchicago.edu/founders/>
- [12] *The Terahertz Wave eBook*. ZOmega Terahertz Corp., 2014. [Online]. Available: [http://dl.z-thz.com/eBook/zomega\\_ebook.pdf.1206.sr.pdf](http://dl.z-thz.com/eBook/zomega_ebook.pdf.1206.sr.pdf). Accessed on: May 19, 2014.
- [13] Philip B. Kurland and Ralph Lerner, eds., *The Founders' Constitution*. Chicago, IL, USA: Univ. of Chicago Press, 1987, Accessed on: Feb. 28, 2010, [Online]. Available: <http://press-pubs.uchicago.edu/founders/>
- [14] J. S. Turner, "New directions in communications," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 1, pp. 11–23, Jan. 1995.
- [15] W. P. Risk, G. S. Kino, and H. J. Shaw, "Fiber-optic frequency shifter using a surface acoustic wave incident at an oblique angle," *Opt. Lett.*, vol. 11, no. 2, pp. 115–117, Feb. 1986.
- [16] P. Kopyt *et al.*, "Electric properties of graphene-based conductive layers from DC up to terahertz range," *IEEE THz Sci. Technol.*, to be published. DOI: 10.1109/TTHZ.2016.2544142.
- [17] PROCESS Corporation, Boston, MA, USA. Intranets: Internet technologies deployed behind the firewall for corporate productivity. Presented at INET96 Annual Meeting. [Online]. Available: <http://home.process.com/Intranets/wp2.htm>
- [18] R. J. Hijmans and J. van Etten, "Raster: Geographic analysis and modeling with raster data," R Package Version 2.0-12, Jan. 12, 2012. [Online]. Available: <http://CRAN.R-project.org/package=raster>



TABLE I  
ADD CAPTION

Dataset	Metric	TEEDAM		SAHP	GRU-CP
		concat	sum		
SO	LL/#events AUROC				
Synthea	LL/#events AUROC				0.85(.014)
ReTweet	LL/#events AUROC	1.83 (0.09) 0.68 (0.03)	1.5 (0.18) 0.67 (0.05)		0.611(0.001)

TABLE II  
ADD CAPTION

Dataset	setting	Model		
		TE	TE+DAM	TE+noise
P12	sc	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)
	mc1	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)
	mc2	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)
p19	sc	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)
	mc1	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)
	mc2	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)

- [19] Teralyzer. Lytera UG, Kirchhain, Germany [Online]. Available: [http://www.lytera.de/Terahertz\\_THz\\_Spectroscopy.php?id=home](http://www.lytera.de/Terahertz_THz_Spectroscopy.php?id=home), Accessed on: Jun. 5, 2014
- [20] U.S. House. 102<sup>nd</sup> Congress, 1<sup>st</sup> Session. (1991, Jan. 11). *H. Con. Res. 1, Sense of the Congress on Approval of Military Action*. [Online]. Available: LEXIS Library: GENFED File: BILLS
- [21] Musical toothbrush with mirror, by L.M.R. Brooks. (1992, May 19). Patent D 326 189 [Online]. Available: NEXIS Library: LEXPAT File: DES
- [22] D. B. Payne and J. R. Stern, "Wavelength-switched passively coupled single-mode optical network," in *Proc. IOOC-ECOC*, Boston, MA, USA, 1985, pp. 585–590.
- [23] D. Ebehard and E. Voges, "Digital single sideband detection for interferometric sensors," presented at the 2<sup>nd</sup> Int. Conf. Optical Fiber Sensors, Stuttgart, Germany, Jan. 2-5, 1984.
- [24] G. Brandli and M. Dick, "Alternating current fed power supply," U.S. Patent 4 084 217, Nov. 4, 1978.
- [25] J. O. Williams, "Narrow-band analyzer," Ph.D. dissertation, Dept. Elect. Eng., Harvard Univ., Cambridge, MA, USA, 1993.
- [26] N. Kawasaki, "Parametric study of thermal and chemical nonequilibrium nozzle flow," M.S. thesis, Dept. Electron. Eng., Osaka Univ., Osaka, Japan, 1993.
- [27] A. Harrison, private communication, May 1995.
- [28] B. Smith, "An approach to graphs of linear forms," unpublished.
- [29] A. Brahm, "Representation error for real numbers in binary computer arithmetic," IEEE Computer Group Repository, Paper R-67-85.
- [30] IEEE Criteria for Class IE Electric Systems, IEEE Standard 308, 1969.
- [31] Letter Symbols for Quantities, ANSI Standard Y10.5-1968.
- [32] R. Fardel, M. Nagel, F. Nuesch, T. Lippert, and A. Wokaun, "Fabrication of organic light emitting diode pixels by laser-assisted forward transfer," *Appl. Phys. Lett.*, vol. 91, no. 6, Aug. 2007, Art. no. 061103.
- [33] J. Zhang and N. Tansu, "Optical gain and laser characteristics of InGaN quantum wells on ternary InGaN substrates," *IEEE Photon. J.*, vol. 5, no. 2, Apr. 2013, Art. no. 2600111
- [34] S. Azodolmolky *et al.*, "Experimental demonstration of an impairment aware network planning and operation tool for transparent/translucent optical networks," *J. Lightw. Technol.*, vol. 29, no. 4, pp. 439–448, Sep. 2011.



**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.

TABLE III  
ADD CAPTION

Dataset	Setting	Center	F1		AUPRC		AUROC	
			DAM	TE+DAM	DAM	TE+DAM	DAM	TE+DAM
P12	sc	1	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)
		2	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)
		3	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)
	mc1	1	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)
		2	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)
		3	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)
	mc2	-	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)
		-	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)
	seft	1	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)
		2	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)
		3	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)
P19	sc	1	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)
		2	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)
		3	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)
	mc1	1	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)
		2	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)
		3	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)
	mc2	-	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)
		-	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)
	seft	-	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)
		-	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)	0.55 (0.02)



**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, Tapei, 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).