ISPT-Net: A Noval Transient Backward-stepping Reduction Policy by Irregular Sequential Prediction Transformer

Yichao Dong[1], Dan Niu[1], Zhou Jin[2], Chuan Zhang[3], Changyin Sun[4] and Zhenya Zhou[5]
1.School of Automation, Southeast University, Nanjing 210096, China
2.Super Scientifc Software Laboratory, China University of Petroleum-Beijing, Beijing, China
3.National Mobile Communications Research Laboratory, Southeast University, Nanjing 211189, China
4.School of Artificial Intelligence, Anhui University, Hefei 230601, China
5.Huada Empyrean Software Co. Ltd, Beijing, China

***Abstract*—In the post-layout simulation for large-scale integrated circuits, transient analysis (TA), determining the time-domain response over a specified time interval, is essential and important. However, it tends to be computationally intensive and quite time-consuming without proper settings of NR initial solution and accurate LTE estimation for determining the next transient timestep, which will lead to a mass of backward-steppings. In this paper, an irregular sequential prediction transformer named ISPT-Net is proposed to predict accurately transient solution as NR initial solution and further obtain precise LTE estimation for setting next timestep. The ISPT-Net is strengthened with timestep positional encoding module (TPE), frequency- and timestep-sensitive muti-head self-attention module (FT-MSA) to enhance irregular sequence feature extraction and prediction accuracy. We assess ISPT-Net in the real large-scale industrial circuits on a commercial SPICE simulator, and achieve a remarkable backward stepping reduction: up to 14.43X for NR nonconvergence case and 4.46X for LTE overlimit case while guaranteeing higher solution accuracy.**

***Index Terms*—Circuit simulation, transient analysis, irregular sequential prediction, deep learning**

I. Introduction

In the post-layout simulation for large-scale integrated circuits, DC analysis, transient analysis (TA) and AC analysis are three fundamental analyses [1]. Among these, P-TA (post-layout transient analysis) is performed to determine the circuit time-domain response and compute the output voltages and currents, as functions of time over a specified time interval $(0, T)$ [2]. It is essential but nearly the most computationally intensive and time-consuming [3]. Besides, P-TA is also required to be repeatedly conducted in some other analyses, e.g. process corner, especially Monte Carlo analysis (usually $\times$ 32,000 times) [4]. Therefore, enhancing the convergence and simulation efficiency of P-TA is quite important and urgent.

In P-TA, two key challenges affect its simulation efficiency. One is how to design an intelligent stepping policy, which achieves a good tradeoff between large step size for high efficiency and small step size for high solution convergence and accuracy [5]. The another obstacle is how to set a good initial solution for Newton-Raphson (NR) method to enhance its convergence and iteration efficiency.

As for the transient stepping policy, apart from the envelope-following (EF) based methods for special fast transient analysis [6], two typical stepping strategies including iteration-count based timestep method (IC-b) and local truncation error (LTE)-based timestep method (LTE-b) are widely used in the commercial SPICE simulators [1][3]. The IC-b method solely relies on the number of NR iterations to decide the step size and the solution accuracy can not be guaranteed. Unnecessarily LTE-based backward steppings usually occur. Therefore, LTE-b method is more widely-used and can generate large enough step size according to the upper limit of next LTE estimation $E^{P}_{\mathrm{T},n+1}$ to obtain high TA efficiency. However, LTE-based and NR-based backward steppings still usually occur due to the inaccuracy estimation of the $E^{P}_{\mathrm{T},n+1}$, and the convergence status lack of history NR iterations [7]. In this case, if the transient solution at the next timestep can be effectively predicted, which can be used to supply a close enough initial solution for NR iterations and also calculate an accurate second order derivative $x''\left(\xi_{n+1}\right)$ for next LTE estimation $E^{P}_{\mathrm{T},n+1}$, then LTE-based and NR-based backward steppings can be largely reduced and high transient efficiency can be obtained.

However, predicting the transient solution of the next timestep is nontrivial and quite challenging, since transient time step varies largely by several orders of magnitude, so it's an irregular time series prediction task. Traditional linear extrapolation exists large prediction error at the nonlinear regions of solution change. Classical RNN/LSTM based deep learning methods [8] do not care for the sample time differences of the input sequence. The NODE (Neural ordinary differential equation) or NCDE methods [9] can solve the irregular time sequence prediction, but the prediction accuracy is not satisfactory for solution curves with large change rate.

In this paper, a transformer-based irregular sequential prediction network named ISPT-Net working seamlessly with timestep positional encoding (TPE) as well as frequency- and

This work was supported by the National Natural Science Foundation of China (Grant No. 62374031), National Key RD Program of China (Grant No. 2022YFB4400400), National Natural Science Foundation of China (Grant No. 62204265, 62234010). (Corresponding author: Dan Niu and Zhou Jin)

timestep-sensitive multi-head self-attention (FT-MSA) modules are proposed to achieve accurate irregular solution prediction, which is utilized to supply a good NR initial solution and accurate LTE estimation for the next timestep. They can largely reduce the backward steppings due to NR nonconvergence and LTE overlimit, and enhance the simulation efficiency of P-TA. The novelty of this work is as follows,

1) To the best of our knowledge, our work is the first DL-enhanced post-layout transient analysis, enabling good initial solution for NR iterations and accurate LTE estimations for balancing the accuracy and convergence, and generating a remarkable backward-stepping reduction in P-TA.

2) An irregular sequential prediction transformer (ISPT-Net) is proposed to predict accurately the transient solution of the next timestep even the timestep varies largely by several orders of magnitude. Considering the irregular sampling, a timestep positional encoding (TPE) module with trigonometric projection and logarithmic mapping is designed to extract and merge the multi-frequency and small-scale temporal features of irregular sequence.

3) A novel frequency- and timestep-sensitive MSA (FT-MSA) module is put forward to learn the time interval and periodical dependencies of transient solution sequence from different representation subspaces, which can enhance the prediction accuracy by pay attention on the nearby transient solutions in the same solution period and the solutions with similar phase in different solution periods.

The ISPT-Net model can be trained by the node solution dataset of pre-layout simulation. It has been implemented in a commerial SPICE-like simulator and verified using the real large-scale industrial circuits. Significant backward-stepping reduction is achieved, i.e., a maximum 14.43X for the NR nonconvergence case and 4.46X for LTE overlimt case.

II. PRELIMINARY

A. Transient simulation

In TA, the time interval $(0, T)$ will be divided into discrete timepoints $(0, t_1, t_2, \cdots, t_n, t_{n+1}, \cdots, T)$ by the timestep control policy [1]. The time step size is $h_n = t_{n+1} - t_n$. At each timepoint, a numerical integration algorithm (e.g. backward Euler, Gear's methods) is employed to replace all the derivatives by finite differences and transform the differential circuit equations into equivalent time-independent nonlinear algebraic equations. The approximation error is the local truncation error (LTE), which is proportional to the timestep h_n. Then the transient timepoint solution is solved by NR iterations. If NR does not converge or the LTE goes beyond the upper error limit, the backward stepping that largely decreases simulation efficiency will be conducted.

B. LTE-based stepping method

Take the widely-used backward Euler intergration method as an example, its LTE is $E_{\mathrm{T},n} = -\frac{{h_n}^2}{2} x''(\xi)$ [1]. h_n is the current time step. Setting the next time step $h_{n+1} = \alpha h_n$, the LTE at the next timepoint is $E_{\mathrm{T},n+1} = -\frac{\alpha^2 h_n^2}{2} x''(\xi_{n+1})$. To meet the requirements $|E_{\mathrm{T},n+1}| \leq e_{\max}$ (error limit) and assuming $x''(\xi_{n+1}) \approx x''(\xi_n)$, then the next step size can be estimated as:

$$h_{n+1} = \alpha h_n \leq \sqrt{\frac{e_{\max}}{|E_{\mathrm{T},n}|}} h_n \tag{1}$$

Although this stepping method is high-efficiency on some circuits, large number of LTE-based and NR-based backward steppings usually occur due to inaccuracy assumption of the derivatives $x''(\xi_{n+1}) \approx x''(\xi_n)$ and lack of considering history NR convergence status. Therefore, if the derivatives $x''(\xi_{n+1})$ can be accurately estimated and a good initial solution for NR iterations can be supplied, backward steppings in TA can be largely reduced.

III. IRREGULAR SEQUENTIAL PREDICTION TRANSFORMER

In this section, an irregular sequential prediction transformer (ISPT-Net) is proposed to predict accurately the transient solution x_{n+1}^P of the next timestep, which is then used as the initial solution of NR iterations and also to further generate accurate LTE estimation $E_{\mathrm{T},n+1}^P$.

A. ISPT-Net Framework

As shown in Fig. 1, the proposed ISPT-Net employs an encoder-decoder transformer structure with stacked encoder layers but only one decoder layer for network lightweight. Our model is to generate the most likely transient solution of next timestep x_{n+1}^P, based on the previous irregular m transient solutions including the current one: $X_n = (x_{n-m+1}, x_{n-m+2}, ..., x_n)$. The model input is the previous solution sequences of the selected "convergence difficulty" nodes, which is easily determined by sorting the backward-stepping nodes in the pre-layout simulation. Usually small number of nodes (<20, as "convergence difficulty" nodes to be predicted) result in the vast majority of backward steppings.

It is known that the typical transformer structure just deals with the time sequence with same time interval since its positional encoding care for the position information of the input sequence [10]. However, the step sizes in transient analysis are not fixed and usually vary by several orders of magnitude [11], [12]. The irregular timestep information must be also encoded and embedded. Therefore, the model input also includes the corresponding m irregular timestep sequence $H_n = (h_{n-m+1}, h_{n-m+2}, ..., h_n^{'})$, where $h_n^{'}$ is obtained by the above-mentioned conventional LTE stepping method (Eq. (1)). Moreover, the timestep sequence H_n is not directly input to the model, but is first encoded by the proposed timestep positional encoding module (TPE), where richer multi-frequence features and enhanced timestep difference features will be extracted. Besides, considering the self-attention mechanism in transformer, the transient time interval matrix $T_{m\times m} = [T_{ij}]$ is also constructed and input to the model, where T_{ij} is the time interval between the transient timepoint T_i and T_j. The model output is the predicted transient solution x_{n+1}^P at the next timestep $h_n^{'}$. As shown in Fig. 1, the extracted irregular

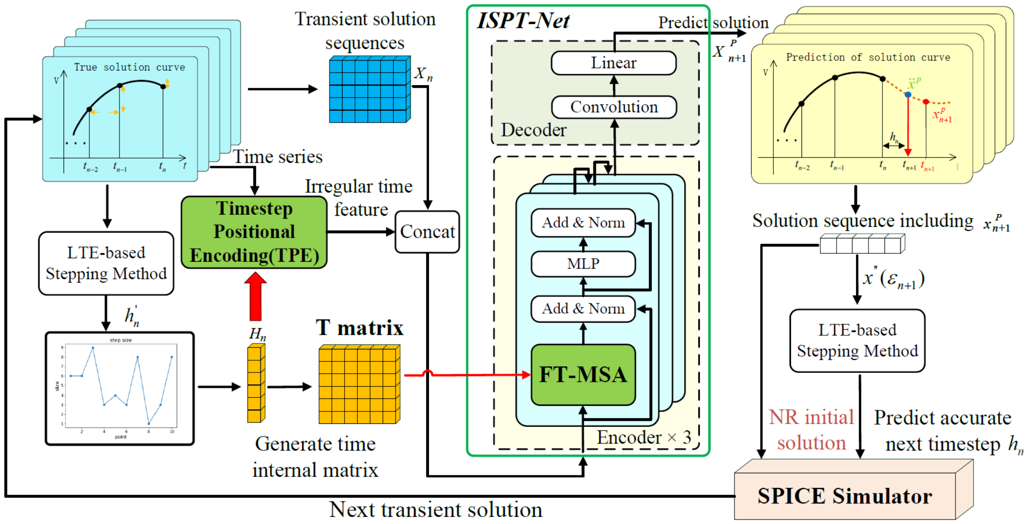

Fig. 1. Proposed irregular sequential prediction transformer for backward-stepping reduction.

timestep features $TPE(H_n)$ will be first concatenated with the solution sequence X_n and then input to the Encoder.

Encoder: The Encoder is composed of a stack of N = 3 identical layers. Each layer is composed of two sub-layers including the proposed frequency- and timestep-sensitive MSA (FT-MSA) and a simple, position-wise multilayer perceptron (MLP). A residual connection is employed around each of the two sub-layers, followed by layer normalization. Different from the typical transformer, FT-MSA has two inputs. The first always is the time interval matrix $T_{m\times m}$, but the second is the concatenated input $ConCat(X_n, TPE(H_n))$ or the output of previous Encoder layer. Considering the characteristics of time sequence prediction task, a mask mechanism is also incorporated in each Encoder layer.

Decoder: Different from the typical transformer with stacked identical layers [10], a lightweight decoder layer is designed to speedup the inference time of prediction network, which consists of a convolution layer and a simple, position-wise fully connected linear feed-forward layer.

B. Timestep Positional Encoding

In the typical transformer, its self-attention module is permutation invariant. Its positional encoding is proposed to combine input embeddings and position information to enable the network to learn the order of the input sequence [10]. However, just order information is not enough in P-TA, since its step sizes are not fixed and vary by several orders of magnitude (usually $10^{-5}s \sim 10^{-14}s$). Not only the order but also the irregular stepsize information must be learned and embedded. As shown in Fig. 2, a timestep positional encoding (TPE) module is proposed to make better use of the order and the temporal information of irregular transient solution sequence. It is expressed by the following equation

$$\mathrm{TPE}\left(t_i, \omega_q\right) = \begin{cases} \sin\left(\omega_q \cdot t_i\right), \\ \cos\left(\omega_q \cdot t_i\right), \\ \alpha \cdot Log\left(t_i - t_{i-1}\right), \\ \beta \cdot t_i \end{cases} \tag{2}$$

where t_i is the ith transient timepoint, $\omega_q (q = 0, 1, 2)$ are three different frequences. α and β are constants. In Eq. (2), trigonometric projection (sine and cosine kernels with different frequencies) is employed alternately to extract richer multi-frequency periodic features of irregular timestep sequence. Besides, the step size $h_i = t_i - t_{i-1}$ in TA is very minute (usually $10^{-5}s \sim 10^{-14}s$) but varies by several orders of magnitude. Thus, the logarithm mapping is utilized to encode and highlight such order-level magnitude differences under the minute time scales. At last, the extracted irregular timestep features will be concatenated with the transient solution sequences X_n as the input of Encoder.

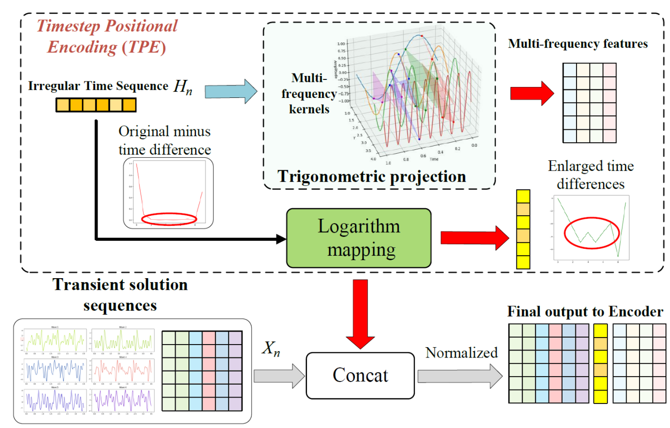

Fig. 2. Proposed timestep positional encoding.

C. Frequency- and Timestep-sensitive MSA

Self-attention (SA) in Transformer has strong ability to capture the long-range dependencies in various series prediction tasks [10]. However, the conventional self-attention models are just sensitive to the relative position relationship of sampled data but not sensitive to the time interval relationship of data. Considering the irregular timesteps in TA, a novel frequency- and timestep-sensitive self-attention (FT-SA) is proposed to enable the prediction model to capture and make better use of the timestep distance and frequency periodic dependencies of irregular solution sequence. Furthermore, FT-SA is extended to frequency- and timestep-sensitive multi-head self-attention (FT-MSA) for enhancing the model to extract the features from different representation subspaces.

As shown in Fig. 3, the proposed FT-SA is represented by the following

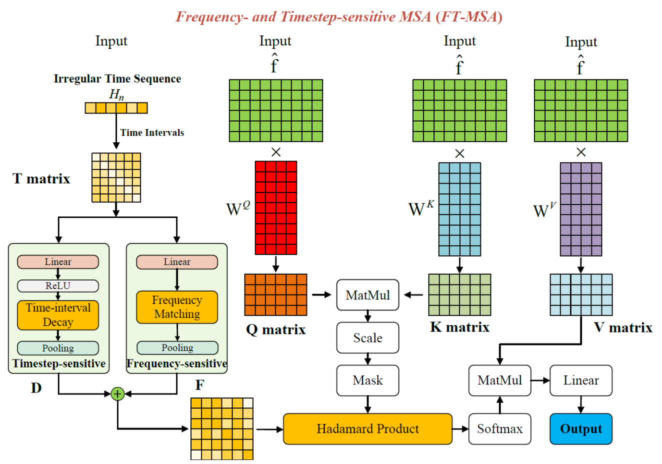

Fig. 3. Proposed frequency- and timestep-sensitive MSA

$$\mathcal{Z}(\mathbf{Q},\mathbf{K},\mathbf{D},\mathbf{F},\mathbf{V}) = \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T \circ (\mathbf{D}+\mathbf{F})}{\sqrt{d_m}}\right)\mathbf{V} \quad (3)$$
$$\mathbf{Q} = \hat{\mathbf{f}}\mathbf{W}^Q, \mathbf{K} = \hat{\mathbf{f}}\mathbf{W}^K, \mathbf{V} = \hat{\mathbf{f}}\mathbf{W}^V$$

where $\hat{\mathbf{f}} \in \mathbf{R}^{n\times d_m}$ is the result of the input embedding. The queries $\mathbf{Q} \in \mathbf{R}^{n\times d_q}$, keys $\mathbf{K} \in \mathbf{R}^{n\times d_k}$ and values $\mathbf{V} \in \mathbf{R}^{n\times d_v}$ are obtained by multiplying $\hat{\mathbf{f}}$ with learnable weight matrics $\mathbf{W}^Q$, $\mathbf{W}^K$ and $\mathbf{W}^V$, respectively. The dimensions of $\mathbf{Q}$, $\mathbf{K}$ and $\mathbf{V}$ are d_q, d_k and d_v. $\circ$ is the Hadamard product. $\mathbf{D}$ and $\mathbf{F}$ are the timestep-sensitive and frequency-sensitive weight matrix, respectively.

As we know, given the vector multiplication $\mathbf{I} = \mathbf{Q}\mathbf{K}^T \in \mathbf{R}^{n\times n}$, where $\mathbf{K}^T$ is the transpose of $\mathbf{K}$, $\mathbf{I}_{i,j}$ can express the influence of input x_j on the input x_i, which is usually sufficient in some prediction tasks, such as machine translation. However, for the transient solution prediction task with irregular timesteps, the self-attention must be modified and the timestep difference and periodic features should be captured and embedded. It has been a prior knowledge that the transient solutions which locate at longer time interval in the same change period or bigger phase difference in different change period from current timepoint, will have weaker impact on the current transient solution. Therefore, the additional timestep-sensitive matrix $\mathbf{D}$ and frequency-sensitive matrix $\mathbf{F}$ are embedded on $\mathbf{I}$ through Hadamard product to enable the self-attention to deal with the irregular transient solution sequence. Then the prediction model can pay more attention on the recent transient solutions in the same period and the transient solutions with the similar phases of different solution periods.

As shown in Fig. 3, $\mathbf{D}$ and $\mathbf{F}$ are obtained from the time interval matrix $\mathbf{T}$ by the proposed timestep-sensitive module and frequency-sensitive module, respectively. The timestep-sensitive module consists of linear layer, ReLU activation function, time-interval decay layer and pooling layer. It is expressed by the following equations

$$\mathbf{Z} = ReLU(W_D \cdot \mathbf{T} + b_D) \quad (4)$$

$$\mathbf{D} = Pooling(\frac{2}{1+e^{\mathbf{Z}}}) \quad (5)$$

In Eq. (4), W_D and b_D are learnable parameters of the linear layer, where W_D is expected to be a positive number when the model is converge. ReLU activation function is employed to ensure that the output element $\mathbf{Z}_{ij}$ is always non-negative. Moreover, exponential decay function (Eq. (5)) is used in the time-interval decay layer. On the one hand, the element value $\mathbf{D}_{ij}$ in attention weight matrix will monotonically decrease if the time interval element $\mathbf{T}_{ij}$ and consequent $\mathbf{Z}_{ij}$ increases, which conforms to our prior knowledge. On the other hand, the weight element value $\mathbf{D}_{ij}$ can be guaranteed ranging between 0 and 1. The pooling layer is used for dimension reduction.

In the frequency-sensitive module, linear layer, frequency matching layer and pooling layer are designed. They are expressed by the following equation

$$\mathbf{F} = Pooling(Sigmoid(\sin{(W_F \cdot \mathbf{T} + b_F)})) \quad (6)$$

where $\mathbf{F}$ is the frequency-sensitive matrix. Note that, W_F as the weight coefficient of the linear layer is equivalent to the angular frequency of the sin function in the frequency matching layer. Different from the fixed frequency feature extraction in TPE (Eq. (2)), the frequency matching layer will learn different angular frequencies to focus on the better matching timepoints. The pooling function aggregates the output weight matrices at different frequencies and maps them to a two-dimensional matrix.

After obtaining the $\mathbf{D}$ and $\mathbf{F}$, the $\mathbf{Q}\mathbf{K}^T \circ (\mathbf{D}+\mathbf{F})$ is divided by $\sqrt{d_m}$ before softmax normalization to enhance the stability of the gradient descent during training.

At last, the multi-head self-attention (MSA) mechanism is employed to increase the representation subspaces, which is expressed by the following equation

$$\mathcal{M}(\mathbf{Q},\mathbf{K},\mathbf{D},\mathbf{F},\mathbf{V}) = Concat(\mathcal{M}_1, ...\mathcal{M}_s)\mathcal{W}^O \quad (7)$$

where $\mathcal{M}_i = \mathcal{Z}(\mathbf{Q}_i,\mathbf{K}_i,\mathbf{D}_i,\mathbf{F}_i,\mathbf{V}_i)$. $\mathcal{W}^O$ is a linear projection matrix.

D. LTE estimation and NR initial solution by ISPT-Net

Based the previous irregular solution sequence X_n and timestep sequence H_n, the ISPT-Net can predict accurately the next transient solution x^P_{n+1}, which can be used as the initial solution for NR iterations directly and also to generate more accurate LTE estimation $E^P_{\mathrm{T},n+1}$ for optimizing the next timestep h_n. The detailed algorithm flow is shown in **Algorithm 1**.

IV. EXPERIMENTS AND RESULTS

A. Network Implementation

We have implemented the ISPT-Net for P-TA in a first-class commercial SPICE simulator (named Com-SPICE), which can supply large-scale real industrial test circuits and cooperative test interface.

For a given circuit, we will accelerate the post-layout transient analysis (P-TA) by the rich information of pre-layout transient analysis (Pre-TA). First, the nodes that cause NR nonconvergence, LTE overlimit and minimum step size in Pre-TA can be obtained. By comparing and sorting the number of backward steppings in Pre-TA, "convergence difficulty" nodes which will be predicted by ISPT-Net in P-TA can be easily selected based on the fact that small number of strong nonlinear nodes (<20) usually result in the vast majority (>70%) of backward steppings. Then the solution and timestep sequences of the "convergence difficulty" nodes in Pre-TA are used to train the ISPT-Net. The training can be finished before the P-TA begins and thus the training time is not important. At last, the trained ISPT-Net can supply accurately next transient solution prediction for NR iterations and LTE estimation since the solution curves of the same node in the P-TA and Pre-TA are almost same but with small time delay. Note that, generally the number of "convergence difficulty" nodes is not large and the prediction time of ISPT-Net (shown in Table IV) is short enough comparing with the time of one NR iteration in large-scale circuits.

B. Backward Stepping Comparison

In this work, the ISPT-Net is proposed to reduce the backward steppings and enhance the simulation efficiency of P-TA. Table I presents the number of total NR iterations and backward steppings of the Com-SPICE with and without the ISPT-Net for 10 industrial circuits with device scale ranging from 3,779 to 1,623,196. The test results of commercial HSPICE are also shown and large number of backward steppings (total-b) also occur ("–" means no results due to simulation error). The

Algorithm 1 LTE estimation and NR initial solution by ISPT-Net

Require: Build algorithm structure:
Input sequence length m, LTE-b stepping method $\mathcal{L}(x)$;
Determine "convergence difficulty" nodes by Pre-TA;
TPE $\mathcal{E}$, Trained ISPT-Net ψ by Pre-TA;
Ensure:
Initial input:
  Solution and stepsize sequence X_n, H_n;
  Get the current timepoint t_n;
while $t_n \neq t_{end}$ **do**
  Get the next time step $h'_n = \mathcal{L}(X_n)$;
  Get the time position encoding $E = \mathcal{E}(X_n, H_n)$;
  Generate T matrix by H_n;
  Predict next transient solution $x^P_{n+1} = \psi(X_n, E, T)$;
  Optimize the time step $h_n = \mathcal{L}(X_n, x^P_{n+1})$;
  Make x^P_{n+1} as initial solution of NR and compute transient solution x_{n+1} at time step h_n;
  $t_n \leftarrow t_n + h_n, n \leftarrow n + 1$;
  Update X_n, H_n;
end while

number of NR iterations and backward steppings (NR-b and LTE-b) are related with the set end time of P-TA, thus their relative reduction ratios by the ISPT-Net are more attractive and also shown. Besides, apart from simulation efficiency, transient solution accuracy should also be guaranteed and is verified by comparing the total LTE (T-LTE).

From this table, it is clear that the ISPT-Net can dramatically reduce the backward steppings of P-TA while keeping higher transient solution accuracy (average 8.11% higher). In detail, the reduction ratios for LTE-based backward steppings (LTE-b) and NR-based backward steppings (NR-b) are 4.46X/14.43X in maximum and 3.07X/1.62X in average, respectively.

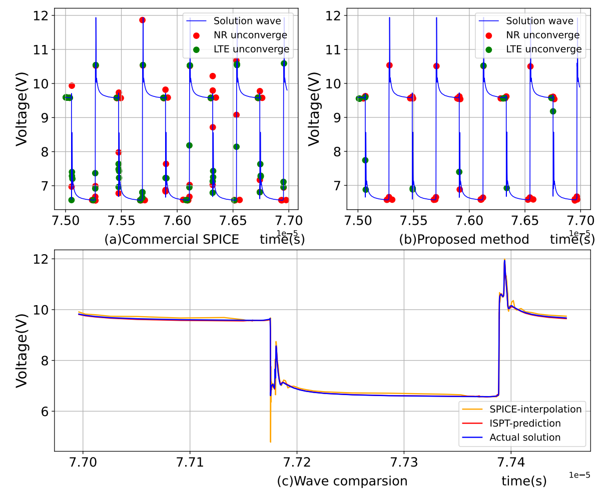

Fig. 4. Transient solution prediction in circuit MP65_post

In order to visualize the prediction performance of the ISPT-Net, Fig. 4 compares the actual transient solution curve and the predicted solution curve of a "convergence difficulty" node in the test case MP65_post. The red dots and blue dots represent NR-based backward-steppings and LTE-based backward-steppings, respectively. From Fig. 4(c), it can be seen that the Com-SPICE generates large prediction error, even up to several volts, where the solution waveform changes dramatically. Compared with it, the predicted solution curve by the ISPT-Net nearly coincides with the actual solution curve. The prediction accuracy can always be maintained within the order of magnitude of 1e-2. Moreover, as shown in Fig. 4(a) and (b), the backward steppings by NR nonconvergence or LTE overlimit are significantly decreased.

In addition, the backward-stepping reduction performance for other "convergence difficulty" nodes in circuit MP65_post is shown in Table II. It can be seen that the backward steppings at nearly all difficult nodes are reduced largely (inf means no backward-stepping). The remaining backward steppings are mainly caused by other unpredicted nodes and they can be reduced further by predicting more nodes.

Furthermore, ablation experiments for TPE and FT-MSA are conducted to verify their effectiveness. Table III gives the results of a example case hed_osc_2. It is clear that both

TABLE I
BACKWARD STEPPING REDUCTION COMPARISONS

Circuits	Device				HSPICE	Com-SPICE				ISPT-Net				Speedup		
	total	mos	r	c	total-b	NR	NR-b	LTE-b	T-LTE	NR	NR-b	LTE-b	T-LTE	NR-b	LTE-b	T-LTE
MP65_post	116963	3687	45104	68145	2767	101963	1299	1101	46.51	85336	702	398	44.23	1.85x	2.77x	4.91%
r3d_post	4483	480	2148	1851	583	29601	372	233	7.17	24253	158	128	6.46	2.35x	1.82x	9.82%
hed_osc_1	3779	296	1981	1486	225	11215	67	183	1.96	9245	18	65	1.69	3.72x	2.82x	13.92%
hed_osc_2	11727	296	6819	4596	223	11771	56	196	2.19	9583	11	64	1.83	5.09x	3.06x	16.40%
hed_trantt	111686	4176	49362	58124	2247	152048	758	1535	131.18	140958	561	605	122.76	1.35x	2.54x	6.42%
dcdc_post	69335	715	50956	17653	3197	83166	1569	1437	39.42	68815	1168	365	36.22	1.34x	3.94x	8.13%
pll_post_rc	409189	8908	139469	260731	-	287253	101	1748	184.65	274111	7	392	163.91	14.43X	4.46X	12.65%
tops_post	1623196	0	308016	1315095	112	173076	0	124	604.37	170326	0	66	551.24	1.00X	1.88X	9.64%
Video_RC	135387	27587	11	106931	-	58950	3	361	84.66	57871	3	156	71.42	1.00X	2.31X	18.54%
pmu_post	308141	4413	298575	4247	-	275707	5796	1382	395.13	240224	3564	468	385.17	1.63X	2.95X	2.59%
Avg	-	-	-	-	-	-	-	-	-	-	-	-	-	1.62x	3.07x	8.11%

#NR-b: number of NR unconvergence #LTE-b: number of LTE overlimit #T-LTE: total sum of the LTE

TABLE II
BACKWARD-STEPPINGS AT OTHER PREDICTED NODES

node	Com-SPICE		ISPT-Net		Speedup	
	NR-b	LTE-b	NR-b	LTE-b	NB-b	LTE-b
520	-	172	-	69	-	2.49x
816	89	101	52	7	1.71x	14.43x
515	88	198	38	9	2.32x	22.00x
522	109	85	-	11	inf	7.73x
514	49	49	13	3	3.77x	16.33x
517	43	52	10	4	4.30x	13.00x
815	164	84	94	10	1.74x	8.40x
802	27	16	11	-	2.45X	inf
513	18	35	-	7	inf	5.00x

TPE and FT-MSA are essential to deal with irregular sequence prediction and reduce the backward steppings.

TABLE III
ABLATION EXPERIMENTS FOR TPE AND FT-MSA

Method	NR	NR-b	LTE-b	T-LTE
Com-SPICE	11771	56	196	2.19
ISPT-Net	9583	11	64	1.83
ISPT-Net w/o TPE	10883	28	102	2.01
ISPT-Net w/o FT-MSA	10170	15	83	1.93

Lastly, considering the computational burden of ISPT-Net, its inference time with different number of predicted nodes in circuit MP65_post is shown in Table IV. It can be seen that the inference time is much less than that of one NR iteration, and it increases slowly even the predicted nodes increase by several times.

TABLE IV
THE INFERENCE TIME UNDER DIFFERENT PREDICTED NODES

Node number	Percentage of unconvergence	Inference time (ms)	Time per NR (ms)
5	45.42%	1.04	2.84
10	63.96%	1.15	
15	68.17%	1.41	
20	71.33%	1.63	
25	73.46%	1.86	

V. CONCLUSIONS

In this paper, we propose a irregular sequential prediction transformer named ISPT-Net to predict accurately next transient solution for NR initial solution and accurate LTE estimations for next timestep, in order to achieve backward-stepping reduction dramatically. The timestep positional encoding (TPE) as well as frequency- and timestep-sensitive muti-head self-attention (FT-MSA) are also designed to deal with irregular sequence and enhance the prediction accuracy. Tests on a commercial SPICE simulator demonstrate a significant backward-stepping reduction while keeping higher solution accuracy.

REFERENCES

[1] T. Nakura, "SPICE Simulation", *Essential Knowledge for Transistor-Level LSI Circuit Design*, Springer, pp. 19-47, 2016.

[2] C. Wang, D. Yang and Q. Chen, "EI-MOR: A Hybrid Exponential Integrator and Model Order Reduction Approach for Transient Power/Ground Network Analysis," in *2022 IEEE/ACM International Conference On Computer Aided Design (ICCAD)*, pp. 1-8, 2022.

[3] K. Kundert, The Designer's Guide to SPICE and SPECTRE®[M]. Springer Science & Business Media, 2006.

[4] M.Firestone, P. Fenner-Crisp, T. Barry, D. Bennett, S. Chang, M. Callahan and S. Knott, "Guiding principles for Monte Carlo analysis", *Environmental Protection Agency*, 1997.

[5] F. Constantinescu, A. G. Gheorghe and M. Nitescu, "A time step choice algorithm for transient analysis of circuits," in *AFRICON 2009*, pp. 1-4, 2009.

[6] D. Linaro, D. d. Giudice, A. Brambilla and F. Bizzarri, "Application of Envelope-Following Techniques to the Simulation of Hybrid Power Systems," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 69, no. 4, pp. 1800-1810, 2022.

[7] G. P. Fang, "A new time-stepping method for circuit simulation," in *2013 50th ACM/EDAC/IEEE Design Automation Conference (DAC)*, pp. 1-10, 2013.

[8] X. Shi, Z. Gao, L. Lausen, H. Wang, D. Yeung, W. Wong and W. Woo, "Deep learning for precipitation nowcasting: A benchmark and a new model," in *30th International Conference on Neural Information Processing Systems (NeurIPS)*, pp. 5617–5627, 2017.

[9] P. Kidger, J. Morrill, J. Foster, and T.Lyons, (2020). "Neural controlled differential equations for irregular time series," in *33th International Conference on Neural Information Processing Systems (NeurIPS)*, pp. 6696–6707, 2020.

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez and I. Polosukhin, "Attention is all you need," in *Advances 30th International Conference on Neural Information Processing Systems (NeurIPS)*, 2017.

[11] Z. Jin, H. Pei, Y. Dong, X. Jin, X. Wu, W. W. Xing, and D. Niu, "Accelerating nonlinear dc circuit simulation with reinforcement learning," in *Proceedings of the 59th ACM/IEEE Design Automation Conference*, pp. 619–624, 2022.

[12] D. Niu, Y. Dong, Z. Jin, C. Zhang, Q. Li and C. Sun, "OSSP-PTA: An Online Stochastic Stepping Policy for PTA on Reinforcement Learning," in *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 42, no. 11, pp. 4310-4323, Nov. 2023.