MISP-Net: Significantly Reducing Transient Backward Steppings via Novel Multi-step Irregular Sequence Prediction

1st Yichao Dong
School of Automation
Southeast University
Nanjing, China
230238503@seu.edu.cn

2nd Dan Niu*
School of Automation
Southeast University
Nanjing, China
danniu1@163.com

3rd Chao Wang
School of Automation
Southeast University
Nanjing, China
230258263@seu.edu.cn

4thZhenya Zhou
Huada Empyrean Software
Beijing, China
2433657050@qq.com

6thChangyin Sun
School of Artificial Intelligence
Anhui University
Hefei, China
3851822122@qq.com

5thZhou Jin
School of integrated Circuits
Zhejiang University
Zhejiang, China
z,jin@zju.edu.cn

***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 time-consuming. Especially, a mass of backward steppings and low simulation efficiency occur without proper settings of Newton-Raphson (NR) initial solution and accurate Local Truncation Error (LTE) estimation. In this work, a novel multi-step irregular sequence prediction model (MISP-Net) is proposed to predict multiple NR initial solutions and precise LTE estimations by just one inference step. This model is constructed by an Irregular Multiple Timesteps Prediction Module (IMTP) and a Irregular Multi-step Solution Prediction Module (IMSP). In IMSP, to improve the irregular prediction performance, a Dual-branch Irregular Feature Pyramid (DIFP) equipped with lightweight Multi-Channel Irregular Time Attention (MITA) are designed. We assess the proposed MISP-Net in the real large-scale industrial circuits on a commercial SPICE simulator. Compared with the commercial SPICE and the SOTA ISPT-Net model, significant backward stepping reductions are achieved: up to 78.57% for NR nonconvergence case and 76.62% for LTE overlimit case, respectively. And the prediction time for NR initial solution in our model is remarkably reduced by up to 5.58× compared to the SOTA ISPT-Net model.**

***Index Terms*—Transient analysis, irregular sequence prediction, feature pyramid, irregular attention mechanism**

I. Introduction

Transient analysis (TA), computing the output voltages and currents over a specific time interval [1], [2], is quite important but usually the most computationally intensive and time-consuming [3], [4] in circuit simulation, especially post-layout transient analysis (P-TA) for large-scale circuits. Moreover, P-TA is usually performed repeatedly in some analyzes, such as Monte Carlo analysis (usually × 32,000 times) [5]. Therefore, accelerating the P-TA efficiency is quite significant and even treated as the bottleneck challenge in some applications (intelligent iterative generation of circuits) [6]–[9].

Currently, two key challenges hinder the P-TA simulation efficiency: one is how to design an intelligent transient stepping policy [10]; another is how to set a good initial solution for Newton-Raphson (NR) iterations and further set an accurate next local truncation error (LTE) estimation $E^{P}_{\mathrm{T},n+1}$. Otherwise, abundant NR nonconvergence-based and LTE overlimit-based transient backward steppings occur, which heavily affect the TA efficiency. In this case, it is clear that if the transient solution at the next timestep can be effectively predicted, it can be used as a close enough initial solution for NR iterations and also to calculate an accurate second order derivative $x''(\xi_{n+1})$ for next LTE estimation $E^{P}_{\mathrm{T},n+1}$ [11], [12]. As a result, LTE-based and NR-based backward steppings can be significantly reduced to achieve high transient efficiency.

However, forecasting accurate transient solution of the next timestep is nontrivial and quite challenging, since it is an irregular time series prediction task. The sizes of transient timesteps can vary significantly by several orders of magnitude ((usually $10^{-5}s \sim 10^{-14}s$)) [13]–[15]. Traditional linear extrapolation methods exist large prediction errors in the nonlinear change regions of solutions [1]. Moreover, well-known deep learning models (e. g. RNN/Transformer) are not applicable for irregular sequence predictions [16]–[18]. The NODE (Neural ordinary differential equations) [19] /NCDE (Neural controlled differential equations) methods [20] can solve the irregular time sequence prediction, but the prediction accuracy is not satisfactory especially for solution curves with large change rates. In [21], ISPT-Net with timestep positional encoding module and multi-head self-attention module is proposed to reduce the TA backward steppings. However, it just achieves one-step solution prediction at each timestep. In this case, the total prediction time by ISPT-Net will be large

978-3-9826741-1-7/DATE26/© 2026 EDAA

enough to affect TA efficiency. Besides, more sophisticated networks (such as Large Models) cannot be adopted due to large prediction inference time.

In this work, a multi-step irregular sequence prediction method named MISP-Net is proposed to achieve more accurate multi-step NR initial solution predictions with low computational overhead and significantly reduce transient backward steppings for high TA efficiency. The main contributions of this work are as follows,

1) A multi-step irregular sequence prediction method (MISP-Net) is proposed, which accurately predicts the multi-step NR initial solutions by a single inference step and significantly reduces the transient backward steppings with low computational overhead. To the best of our knowledge, it is the first time to achieve multi-step irregular solution predictions by deep learning model.
2) To predict the future multiple irregular timesteps, an Irregular Multiple Timesteps Prediction Module (IMTP) is proposed, which breaks the key bottleneck for the application of sophisticated prediction models (such as Large Models).
3) To enhance the prediction accuracy for irregular time series, a Dual-branch Irregular Feature Pyramid (DIFP) is designed to model the variation and amplitude features of solution curves, respectively. By aggregating multi-scale features of solution curves through the pyramid branches, the ability of modeling on different sample granularities and irregular series prediction are greatly improved.
4) Considering the circuit physical constraints, a novel lightweight Multi-channel Irregular Timestep Attention Module (MITA) is designed to capture spatiotemporal correlation features between different solution curves at irregular timesteps, therefore enhancing the prediction accuracy of solution curves.

Our method has been implemented in a commerial SPICE-like simulator and verified using the real large-scale industrial circuits.

II. Background

A. Transient Simulation

In TA, the simulation time $(0, T)$ will be segmented into a sequence of distinct time intervals $(0, t_1, t_2, \ldots, t_n, t_{n+1}, \ldots, T)$ by the transient timestep control policy [1]. The timestep size is defined as $h_n = t_{n+1} - t_n$. At each timepoint, a numerical integration algorithm (e.g. backward Euler) is used to substitute the derivatives with finite difference approximations [3]. The h_n is usually calculated by the estimated LTE. Then the next transient solution is computed by the NR iterations. If the NR iterations fail to converge or the post-verified LTEs exceed the accuracy threshold, the transient backward steppings occur and simulation efficiency is significantly decreased .

B. LTE-based Stepping Method

The LTE-based stepping method is widely used in commercial SPICE simulators. Taking the backward Euler integration as an example and assuming the next timestep is $h_{n+1} = \alpha h_n$ (h_n is the current time step), the LTE at the next timepoint can be calculated as $E_{\mathrm{T},n+1} = -\frac{\alpha^2 h_n^2}{2} x''(\xi_{n+1})$ ($x''(\xi_{n+1})$ is the second order derivative of solution curve). To meet the solution error limit $|E_{\mathrm{T},n+1}| \leq e_{\max}$ and assuming $x''(\xi_{n+1}) \approx x''(\xi_n)$, then the next timestep size can be estimated as $h_{n+1} = \alpha h_n \leq \sqrt{\frac{e_{\max}}{|E_{\mathrm{T},n}|}} h_n$. If the assumption $x''(\xi_{n+1}) \approx x''(\xi_n)$ has big error, the estimated next LTE $|E_{\mathrm{T},n+1}|$ and the resulting h_{n+1} will have big error. Then the LTE-based backward stepping will occur. Therefore, it is clear that that if the next transient solutions can be accurately predicted, then the accurate $x''(\xi_{n+1})$ can be easily obtained. In conclusion, the accurate predicted next transient solutions are vital for the estimations of NR initial solution and $x''(\xi_{n+1})$, and further for decreasing the two transient backward steppings [11], [12].

III. Multi-step Irregular Sequence Prediction

In this section, a novel multi-step irregular sequential prediction framework (MISP-Net) is proposed to generate the most likely multi-step (k-step) future transient solution predictions, S^P_{n+1}, S^P_{n+2}, ..., S^P_{n+k}, based on the previous J transient solutions including the current one: S_{n-J+1}, S_{n-J+2}, ..., S_n. Then the k-step future transient solution predictions S^P are used to fit the future transient solution curves, which can interpolate to obtain the future multi-step (ususlly equal to k) accurate initial solutions for NR iterations and also further generate accurate multi-step LTE estimations E^P_T. It means that just approximate $1/k$ of model inference time is taken for obtaining each accurate NR initial solution prediction. It breaks the inference time limit upon those excellent but non-lightweight models for irregular node solution predictions (such as employing Large Models to predict hundreds of node solutions becomes possible).

A. MISP-Net Prediction Framework

To achieve accurate multi-step future transient solution predictions, **two coupled challenges must be solved**: One is how to predict future multiple irregular timesteps? Another is how to predict future multiple irregular transient solutions? In this work, the irregular timestep and solution prediction will be decoupled to decrease the error accumulation and achieve high prediction accuracy. as well as better generalization.

As shown in Fig. 1, the historical irregular sequences X_h (timestep sequence, solution sequences and some critical supplementary features) are input to the proposed MISP-Net (including irregular multiple timesteps prediction module (IMTP) and irregular multi-step solution prediction module (IMSP)) to obtain a predicted interpolation curve function F_{ic} (valid in the predicted time interval t_k (=h^P_{n+1} +...+h^P_{n+k})). Subsequently, the SPICE simulator will generate multiple **future timesteps** (h'_n) based on the transient stepping policy, and iteratively invoke the F_{ic} to derive the accurate **NR initial solution** S'_n **and LTE estimations** E'_T, which can significantly reduce the transient backward steppings. Note that, the MISP-Net will not be conducted until the simulation time t exceeds

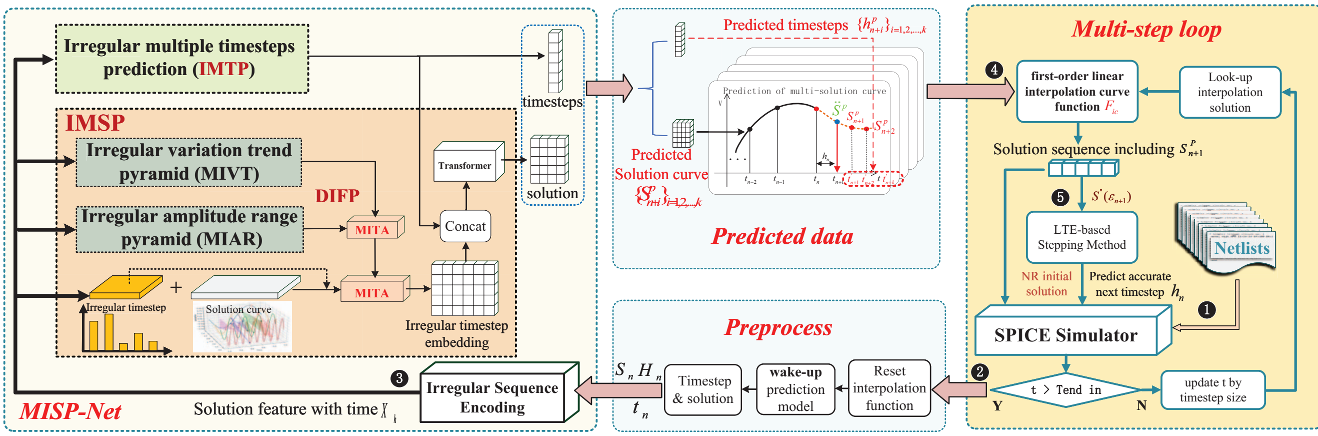

Fig. 1. Proposed multi-step irregular sequence prediction framework (MISP-Net).

the t_k. Therefore, the prediction time for each timestep NR initial solution by model is largely reduced.

B. Irregular Multiple Timesteps Prediction Module (IMTP)

how to predict future multiple irregular timesteps? First, what inputs are required? Apart from the history irregular timestep sequence, history transient solution sequences are also required, since the solution changes seriously affect the irregular timestep prediction [14], [22]. Note that the transient step sizes are not fixed and vary by several orders of magnitude (usually $10^{-5}s \sim 10^{-14}s$), logarithmic mapping $LM(h) = Log(h)$ is designed to encode and highlight such order-level magnitude differences. Moreover, to extract richer multi-frequency periodic features of irregular timestep sequences, we design the trigonometric projection $TR(t) = \{sin(w_q t), cos(w_q t)\}$ (t is transient timepoint, $w_q (q = 0, 1, 2)$ are three different frequencies). Then the total input sequence data X_h is formulated as $X_h = Encoding(S_n, t, h) = \{S_n, TR(t), \alpha \cdot LM(h), \beta \cdot t\} \in \mathbf{R}^{l \times J}$. α and β are constants, $l = M + 8$. M is the set node number of predicted transient solutions. J is the set number of history transient solutions. Next, the two-stage irregular multiple timesteps prediction module (IMTP) is proposed to accurately predict future multiple transient timesteps.

It is known that the GRU (Gated Recurrent Unit) model [23] is effective to handle the time series prediction due to simple structure and low computational complexity. Considering the model inference time affects the TA efficiency, the GRU framework is selected to predict the timesteps with minimal computational overhead. However, two key difficulties arise due to GRU intrinsic recursive structure: 1) how to solve the input dimension mismatch? since only timesteps are predicted here and the corresponding solution embeddings are lost; 2) how to accurately map the GRU output (0 to 1) to timestep? The large timestep variations in the order of magnitude must be considered. For difficulty 1, we design a two-stage prediction framework, including input embedding GRU (GRU-e) and timestep prediction GRU (GRU-p).

Stage 1: input embedding GRU (GRU-e) In this stage, the history sequence data X_h is input to the GRU-e model to extract the irregular timestep features and compress them into a one-dimensional vector τ_n, as shown in Fig. 2. Then the vector τ_n with irregular features is input to the GRU-p as the initial state. The input of GRU-p is not X_h but just the predicted timestep h_n^P, which solves the dimension mismatch of recursive structure.

Stage 2: timestep prediction GRU (GRU-p) As shown in Fig. 2, by using τ_n as initial state and the current step size h_n as input [17], the GRU-p is conducted to obtain the next irregular time feature state τ_{n+1}. Next, difficulty 2: how to map the state output τ_{n+1} to timestep prediction h_{n+1}^P is considered. Due to the large variations in the order of timestep magnitude (usually $10^{-5}s \sim 10^{-14}s$), linear mapping the GRU output will result in poor accuracy for those timesteps with small magnitudes. In this work, a sophisticated probabilistic model is utilized to characterize the statistical properties of the stepsizes. In detail, we employ a log-normal mixture function [24] to model the probability density function, which is defined as follows:

$$p(h \mid \boldsymbol{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}) = \sum_{k=1}^{K} w_k \frac{1}{h \boldsymbol{\sigma}_k \sqrt{2\pi}} \exp\left(-\frac{(\log h - \mu_k)^2}{2\boldsymbol{\sigma}_k^2}\right), \quad (1)$$

where $\boldsymbol{w}$ represents the weights of the mixture densities, $\boldsymbol{\mu}$ denotes the means of the mixture densities, and $\boldsymbol{\sigma}$ signifies the variances of the mixture densities. The three parameters are obtained from $\tau_{n+i} (i = 1, ...k)$ through different linear mapping relationships, respectively.

To enable the computation of gradients for the sampling results, the following formulas are designed:

$$\begin{cases} \boldsymbol{z} = \text{softmax}\left(\frac{\log \boldsymbol{w} + \boldsymbol{o}}{\zeta}\right), \\ \varepsilon \sim \text{Normal}\,(0, 1)\,, \\ \tau = \exp\left(\boldsymbol{\sigma}^T \boldsymbol{z} \cdot \varepsilon + \boldsymbol{\mu}^T \boldsymbol{z} \cdot \varepsilon\right), \end{cases} \quad (2)$$

where o is sampled from the independent and identically distributed standard Gumbel distribution, ζ is a parameter. As ζ approaches zero, samples from the Gumbel-softmax distribution approach exact samples from the discrete distribution [25]. The different kernels of the log-normal mixture distribution can handle the stepsizes under different magnitudes separately, thereby enhancing the timestep prediction accuracy.

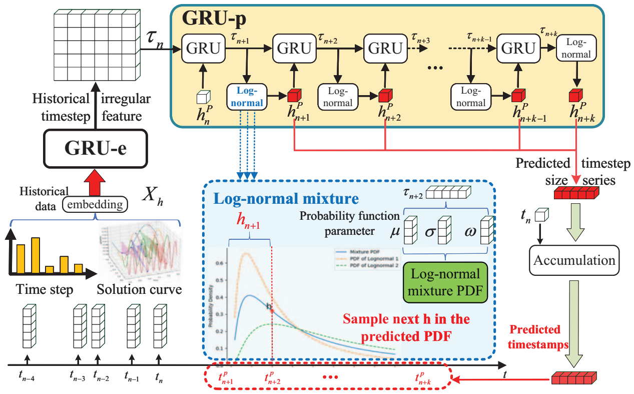

Fig. 2. Proposed irregular multiple timesteps prediction module (IMTP)

Meanwhile, due to the randomness of timestep sampling, the training of the solution prediction model will have better generalization performance.

Once the probability density function is obtained, the predicted timestep h_{n+1}^P is determined by sampling. Together with τ_{n+1} , it is then fed into the next computational unit of GRU-p, which gives the prediction for the subsequent timestep. This iterative process is repeated k times. As a result, we achieve an accurate predicted timestep sequence h_n^P of length k. The multiple irregular predicted timestep sequence h_n^P and X_h will be input to the following irregular multi-step solution prediction module (IMSP) to obtain an accurate interpolation curve function prediction F_{ic} for SPICE simulator.

C. Irregular Multi-step Solution Prediction Module (IMSP)

As shown in Fig. 1, IMSP equipped with Dual-branch Irregular Feature Pyramid (DIFP) and Multi-channel Irregular Timestep Attention Module (MITA) is proposed to accurately predict the multi-step irregular transient solutions.

How many node solutions are required to be predicted? It is found that usually a small number of nodes (<25, only "convergence difficulty" nodes to be predicted) even in large-scale circuits generate the vast majority of transient backward steppings. Thus, the model inputs just employ the history transient solution sequences of the selected "convergence difficulty" nodes, which are easily determined by sorting the backward stepping nodes in the pre-layout simulation.

How to encode the irregular solution sequence features? For the irregular timestep sequences, logarithm mapping $LM(h)$ and trigonometric projection $TR(t)$ have been proposed to deal with the order-level "magnitude differences" features and extract multi-frequency "periodic" features, respectively. Moreover, circuit irregular transient solutions are correlated with varying amplitude range and periodic trend features on different sample scales/granularities. In this work, dual-branch irregular feature pyramids including multi-granularity irregular variation trend pyramid (MIVT) and multi-granularity irregular amplitude Range pyramid (MIAR) are proposed to extract coarse-to-fine solution features.

Branch 1: Multi-granularity Irregular Variation Trend Pyramid (MIVT) The transient convergence and transient timesteps are strongly associated with solution variation slopes, which are also critical to predict the transient solutions under the irregular timesteps. To extract the local and global variation trend features, the least squares method with second-order norm optimization is the best linear-fitting unbiased estimator (under the Gauss-Markov theorem). It is utilized to fit and calculate the trend/slope features v_g on different sample granularity set Π_g for $g = 1, 4, 16, 64$ (pyramid structure). The formula is $v_g = \frac{g\sum h_n S_n - \sum h_n \sum S_n}{g\sum h_n^2 - (\sum h_n)^2}$, where the summation symbol $\sum$ represents the aggregate over all transient solutions at each respective granularity g. The least squares fitting has an accurate and low-computational cost process. It also shows good robustness against data noise.

Branch 2: Multi-granularity Irregular Amplitude Range Pyramid (MIAR) In this work, a temporal convolution Conv-t $(S_n,h_n)= \sum wS_nh_n$ is designed to extract the comprehensive amplitude range features across the different solution granularities. w is the weight of the convolution kernel. h_n is added to retain the irregular timestep features in the deep neural network. The temporal convolution can be regarded as a low-pass filter to reflect the steady-state amplitude range of irregular transient solutions.

How to fuse the irregular solution sequence features from different granularities? In order to improve the prediction performance, various attention modules are often added. However, conventional attention modules without temporal information are ill-equipped to adaptively capture and fuse the critical irregular solution features. In this work, a multi-channel irregular time attention module (MITA) with three branches is proposed to effectively extract and fuse features from irregular time series.

The first branch encapsulates global contextual features by global average pooling, while the third branch captures local contextual features. Both branches leverage one-dimensional convolution kernels to synthesize the interrelationships between channels (different nodes), while the attention module remains as lightweight as feasible [26]; The second branch includes the designed irregular timestep attention module (ITA) and an exponential decay function $\mathbf{D} = \frac{2}{1+e^{-\mathbf{z}}}$. z is the output of ITA module. It adaptively adjusts the attention to focus on the solutions with large variation trends or those solutions under small and dense timesteps. For the input variable $\mathbf{X}$, and timestep series $\mathbf{H}$, the attention formulas are written as:

$$\begin{cases} \mathbf{L}(\mathbf{X}) = \mathcal{B}\left(\mathrm{Conv}_2\left(\delta\left(\mathcal{B}\left(\mathrm{Conv}_1(\mathbf{X})\right)\right)\right)\right) \\ \mathbf{G}(\mathbf{X}) = \mathcal{B}\left(\mathrm{Conv}_4\left(\delta\left(\mathcal{B}\left(\mathrm{Conv}_3(GAP(\mathbf{X}))\right)\right)\right)\right) \\ T(\mathbf{X}) = \mathcal{B}(\mathrm{Decay}(\mathrm{ITA}(\mathbf{H}))) \\ \mathbf{X}' = \mathbf{X} \otimes \mathbf{M}(\mathbf{X}) = \mathbf{X} \otimes \sigma(\mathbf{L}(\mathbf{X}) \oplus \mathbf{G}(\mathbf{X}) \oplus \mathbf{T}(\mathbf{X})) \end{cases} \quad (3)$$

where $\mathrm{Conv}_i(i = 1, 2, 3, 4)$ is point-wise convolution, $\mathcal{B}$ is batchnorm function, δ is activate function, GAP is the GlobalAvgPooling.

Considering the ITA module, the irregular timestep density in each granularity $d_g = \frac{\sum h_n}{g}$ is defined to represent the timestep distribution, which affects the solution prediction accuracy. As show in Fig. 1, irregular variation trends under different granularity are extracted in MIVT module and fused

TABLE I
BACKWARD STEPPING REDUCTION AND MODEL INFERENCE TIME COMPARISONS WITH COMMERCIAL SPICE AND ISPT-NET

Circuits	device	C-SPICE			C-SPICE with ISPT-Net				C-SPICE with MISP-Net				Speed-up (vs. C-SPICE)		
		NR-b	LTE-b	total-t	NR-b	LTE-b	total-t	infer-t	NR-b	LTE-b	total-t	infer-t	NR-b	LTE-b	infer-t
MP6517	8297	1299	1101	480.5	902	398	417.2	85.1	877	430	366.3	21.1	32.49%	60.94%	4.03×
r3d_post	4483	372	233	117.4	158	128	99.5	25.9	126	111	76.0	6.1	66.13%	52.36%	4.27×
hed_osc_reduced	3779	67	183	42.3	18	65	36.5	10.1	16	53	30.2	2.5	76.12%	71.04%	4.03×
hed_osc_post	11727	56	196	56.6	11	64	50.9	10.2	12	50	40.9	2.5	78.57%	74.49%	4.15×
hed_trantt	111686	758	1535	3114.6	561	605	2772.8	131.7	513	572	2575.8	30.5	32.32%	62.74%	4.31×
dcdc_post_spice	69335	1569	1437	376.6	1168	365	330.1	77.4	981	336	253.8	17.3	37.48%	76.62%	4.48×
pmu_post	308141	5606	1222	7420.5	3227	419	6877.3	204.8	2715	329	6585.5	53.1	51.57%	73.08%	3.86×
starc	12424	87	32409	31344.4	123	21069	30249.6	1134.3	46	18324	25888.7	240.2	47.06%	43.46%	4.72×
mp86901cr3	16565	2850	221	1711.7	1506	123	1490.0	133.4	1311	108	1418.8	34.2	54.00%	51.13%	3.90×
mtdialog	11854	399	0	200.8	197	0	181.3	8.6	115	0	154.1	2.0	71.18%	-	4.42×
zip	1047182	3855	242	25450.3	2019	174	23705.5	323.9	1613	144	22941.3	74.0	58.16%	40.50%	4.38×
mp86881	430016	2668	351	1371.5	1963	279	1344.4	85.0	1543	258	1242.1	19.6	42.17%	26.50%	4.34×
speed_beh	4106	171	2083	532.1	163	1226	514.7	43.6	121	870	434.7	7.8	29.21%	58.23%	5.58×
mp4653	30191	1471	316	792.0	1039	227	793.0	63.6	889	207	709.5	15.1	39.56%	34.49%	4.20×
mp1470	3144	4487	7158	772.5	3897	5013	1073.8	354.9	3151	3911	753.5	86.0	29.77%	45.36%	4.13×
Avg	-	-	-	-	-	-	-	-	-	-	-	-	49.72%	55.07%	4.32×

#NR-b & #LTE-b: number of NR nonconvergence and LTE overlimit   #total-t: total time consumption (s)   #infer-t: model inference time consumption (s)

by the MIAR module, where the attention is required to focus on the irregular solution predictions with high variation slopes and high timestep densities. In this case, $ITA(I, h_n)$ is designed as $ITA(I, h_n) = \frac{\sum|\delta I * h_n|}{d_g}$, where δI is the input deviation of ITA module. Moreover, attention in MIAR module focuses on the feature fuse with high timestep densities and thus $ITA(I, h_n) = d_g$ is given.

Multi-step solution prediction: The output of the DIFP is fused with the original data X_h using MITA to obtain historical irregular timestep embedding. This embedding is then concatenated with the multiple timesteps $\{h^P_{n+i}\}_{i=1,2,\ldots k}$ output by the IMTP model and fed into a standard Transformer [18], [27] model, which ultimately produces the predicted solution S^P_{n+i} for each timepoint.

IV. EXPERIMENTS AND RESULTS

A. Network Training

The proposed MISP-Net has been implemented for post-layout transient analysis (P-TA) in a first-class commercial SPICE simulator (named C-SPICE), which supplies large-scale real industrial test circuits and test interface.

In this work, the post-layout transient analysis will be accelerated using the rich timestep and solution information in pre-layout transient analysis (Pre-TA). In Pre-TA, the nodes that cause NR nonconvergence, LTE overlimit and minimum step size can be selected. It is found that small number of strong nonlinear nodes (<25) usually result in the vast majority (>70%) of transient backward steppings. By comparing and sorting the number of node backward steppings, "convergence difficulty" nodes predicted in P-TA can be easily determined. Then we train the proposed MISP-Net using the solution and timestep sequences of the "convergence difficulty" nodes in Pre-TA, since the variation trends of solution curves of the same nodes in the P-TA and Pre-TA are similar. And the training can be easily finished before the P-TA begins. Moreover, our prediction achieve multi-step sequence prediction and the inference time for one NR initial solution prediction is largely decreased, which can nearly be neglected compared with one NR iteration time in large-scale circuits. Our training device is NVIDIA RTX3090 24GB GPU. The detailed architecture of the MISP-Net model is as follows:

1) IMTP: Both GRU-e and GRU-p use 64 hidden neurons, with input dimensions of 33 and 1, respectively.
2) MITA: The number of hidden neurons is 64, and the number of point-wise convolutional kernels is 4, the number of log-normal mixture distribution kernel is 64.
3) DTFP: The input dimension is 33, the number of internal hidden neurons is 256, the number of attention stacking layers is 2, and the pyramid is divided into four layers with granularities of 1, 4, 16, and 64, respectively. The size of the convolutional kernels in the temporal convolution is 3×3. The standard transformer model has 256 hidden neurons and 3 attention stacking layers.

B. Backward Stepping Comparisons

In this work, the MISP-Net is proposed to reduce the backward steppings and accelerate the time-consuming P-TA simulation. To evaluate the performance, 15 industrial circuits with device scale ranging from 3,144 to 1,047,182 are tested. Table I shows the number of backward steppings, total simulation time and model inference time by the advanced comercial SPICE, the SOTA ISPT-Net and the proposed MISP-Net. Note that, the relative reduction ratios of two backward steppings (NR-b and LTE-b) are also shown, and they are more important, since the reduction number can be largely enhanced when the set transient end time is increased or the P-TA are repeatedly conducted (such as Monte Carlo analysis).

From this table, it is clear that compared to the advanced commercial SPICE simulator (C-SPICE, using improved extrapolation method), the NR-based and LTE-based backward steppings under the proposed MISP-Net can be significantly reduced by average 47.26% and 51.87%, maximum 76.12% and 76.62%, respectively. Thus, the total time consumption (total-t, including model inference time and circuit simulation time) is largely reduced, and this efficiency enhancement can be dramatically enlarged, since transient analysis is usually conducted repeatedly thousands of times.

In addition, compared to the SOTA ISPT-Net, the MISP-Net can also achieve better backward stepping reduction performance due to higher solution prediction accuracy. More importantly, the inference time taken for one NR initial prediction is remarkably reduced (average 4.32×). The proposed multi-step prediction framework breaks the inference time limit and makes more accurate but complicated models even Large Models for more node predictions become possible.

C. Typical Circuit Analysis

In order to present and visualize the prediction performance of the proposed MISP-Net, the backward steppings in the test case MP6517 are given in detail. In this case, just 13 nodes can cover more than 74.83% of the NR-based backward steppings and 82.65% of the LTE-based backward steppings. Thus, just predicting small number node solutions ("convergence difficulty") can effectively reduce transient backward steppings. Other nodes still use the linear extrapolation method.

The backward stepping reduction ratios at each "convergence difficulty" node of the MP6517 circuit are also tested. The average backward steppings can be decreased by 86.72% for LTE-b and 85.83% for NR-b. It means that the prediction accuracy is high and the backward steppings at "convergence difficulty" are almost vanished.

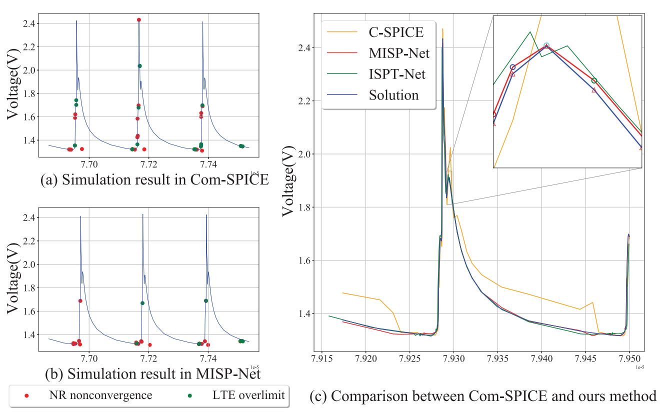

Fig. 3. Transient solution prediction in industrial circuit MP6517

In addition, the actual transient solution curve and the predicted solution curves of a "convergence difficulty" node by the C-SPICE (improved extrapolation), ISPT-Net and MISP-Net are compared in Fig. 3. The red dots and blue dots denote the NR-based backward steppings and LTE-based backward steppings, respectively. From Fig. 3(c), it is clear that the C-SPICE generates large prediction errors at where the solution curve changes dramatically. Compared with it, the predicted solution curves by the ISPT-Net and the proposed method nearly coincide with the actual solution curve. However, our method can achieve more prediction accuracy seen from the enlarged small window. Our prediction accuracy can always be maintained within the magnitude of 0.02V by assuming the prediction time for each NR initial solution just 0.35 ms (but each NR iteration time is 3.1 ms). Moreover, from Fig. 3 (a) and (b), the two backward steppings are significantly decreased.

TABLE II
BACKWARD STEPPING REDUCTION WITHOUT DIFFERENT MODULES

Circuits	Ours	w/o DIFP	w/o IMTP	w/o MITA
	total-b	total-b	total-b	total-b
MP6517	**1187**	1732	1347	1508
r3d_post	**237**	302	320	542
hed_osc_reduced	**69**	108	**69**	162
hed_osc_post	**62**	110	108	271
hed_trantt	**1085**	1847	1148	1543
dcdc_post_spice	1317	1681	**1260**	1614
pmu_post	**3044**	4125	3241	4471
starc	**18370**	22925	18502	19579
mp86901cr3	**1419**	1846	1501	2938
mtdialog	**117**	178	134	188
zip	**1757**	2258	2138	4417
mp86881	1801	2266	**1726**	2381
speed_beh	**991**	1064	1079	2360
mp4653	**1096**	1270	1287	1436
mp1470	7062	8517	**6856**	9258

#total-b: number of NR nonconvergence and LTE overlimit

D. Ablation Experiments

Furthermore, ablation experiments for DIFP (Dual-branch pyramid, MIVT+MIAR), IMTP (comparing with simple GRU prediction) and MITA (with/without ITA) are conducted to verify their effectiveness, as shown in Table II. It is clear that both variation trend and amplitude range pyramid for irregular sequence feature extraction, probabilistic model for irregular timestep output, as well as irregular timestep attention module (ITA) for multi-granularity feature fusion, are essential for accurate multi-step irregular sequence prediction and significantly reducing the transient backward steppings.

V. CONCLUSIONS

In this paper, a high-performance transient backward stepping reduction framework (MISP-Net) is proposed to predict multiple accurate initial solutions for NR iterations and precise LTE estimates for subsequent timesteps with low computational overhead. In MISP-Net, to improve the prediction performance of the irregular solution curves, a Dual-branch Architecture of Irregular Feature Pyramid (DIFP) equipped with lightweight Multi-Channel Irregular Time Attention (MITA) are designed. The proposed multi-step prediction framework using the Irregular Multiple Timesteps Prediction Module (IMTP) breaks the inference time limit and makes more accurate but complicated models even Large Models for more node predictions become possible. We assess the proposed method in the real large-scale industrial circuits. Our method achieves significant backward stepping reduction (up to 78.57% for NR nonconvergence case and 76.62% for LTE overlimit case, respectively) by a remarkably reduction for the prediction time of NR initial solution compared to the SOTA ISPT-Net model.

VI. ACKNOWLEADGEMENT

Corresponding author is Dan Niu. This work is partially supported by NSFC (Grant No. 62374031), Natural Science Foundation of Jiangsu Province (BK20240173), NSFC (Grant No. 62204265, 62234010, 92473107, 62331009)

References

[1] T. Nakura, *SPICE Simulation*, pp. 19–47. Singapore: Springer Singapore, 2016.

[2] Q. Chen, "Ei-nk: A robust exponential integrator method with singularity removal and newton-raphson iterations for transient nonlinear circuit simulation," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 41, pp. 1693–1703, June 2022.

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

[4] D. Niu, Y. Tao, Z. Jin, Y. Dong, C. Wang, and C. Sun, "Islu: Indexing-efficient sparse lu factorization for circuit simulation on gpus," in *Proceedings of the 43rd IEEE/ACM International Conference on Computer-Aided Design*, ICCAD '24, (New York, NY, USA), Association for Computing Machinery, 2025.

[5] M. Firestone, P. Fenner-Crisp, T. Barry, D. Bennett, S. Chang, M. Callahan, and S. Knott, "Guiding principles for monte carlo analysis," tech. rep., Environmental Protection Agency, 1997.

[6] 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," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 42, pp. 4310–4323, Nov. 2023.

[7] Y. Dong, D. Niu, Z. Jin, C. Zhang, Q. Li, and C. Sun, "Adaptive stepping pta for dc analysis based on reinforcement learning," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 70, pp. 266–270, Jan. 2023.

[8] Z. Jin, W. Li, H. Pei, X. Zha, Y. Dong, X. Jin, X. Wu, D. Niu, and W. W. Xing, "Ml-pta: A two-stage ml-enhanced framework for accelerating nonlinear dc circuit simulation with pseudo-transient analysis," *IEEE Transactions on Computers*, vol. 74, no. 10, pp. 3319–3331, 2025.

[9] T. Zhao, Z. Yin, H. An, X. Yang, Z. Jin, J. Shen, and H. Xu, "pdgrass: A fast parallel density-aware algorithm for graph spectral sparsification," in *2025 IEEE High Performance Extreme Computing Conference (HPEC)*, pp. 1–7, 2025.

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

[11] I. N. Hajj, "Circuit theory in circuit simulation," *IEEE Circuits and Systems Magazine*, vol. 16, no. 2, pp. 6–10, 2016.

[12] S. Hamedi-Hagh, *Computational Electronic Circuits: Simulation and Analysis with MATLAB®*. Springer International Publishing, 2022.

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

[14] R. Janssens, L. Lapon, V. Naessens, and B. De Decker, "A review on irregular time series data modeling with recurrent neural networks," *Neurocomputing*, vol. 512, pp. 235–255, 2023.

[15] 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*, DAC '22, (New York, NY, USA), p. 619–624, Association for Computing Machinery, 2022.

[16] 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.

[17] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[18] M. Jin, S. Wang, L. Ma, Z. Chu, J. Y. Zhang, X. Shi, P.-Y. Chen, Y. Liang, Y.-F. Li, S. Pan, and Q. Wen, "Time-LLM: Time series forecasting by reprogramming large language models," in *The Twelfth International Conference on Learning Representations (ICLR)*, 2024.

[19] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, "Neural ordinary differential equations," in *Advances in Neural Information Processing Systems*, vol. 31, 2018.

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

[21] Y. Dong, D. Niu, Z. Jin, C. Zhang, C. Sun, and Z. Zhou, "Isptnet: A novel transient backward-stepping reduction policy by irregular sequential prediction transformer," in *2024 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pp. 1–6, IEEE, 2024.

[22] O. Shchur, A. C. Türkmen, T. Januschowski, and S. Günnemann, "A survey on neural temporal point processes," *IEEE Transactions on Knowledge and Data Engineering*, 2023.

[23] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734, Association for Computational Linguistics, 2014.

[24] O. Shchur, M. Bilos, and S. Gunemann, "Intensity-free learning of temporal point processes," in *Proceedings of the 37th International Conference on Machine Learning (ICML)*, (Vienna, Austria), pp. 1234–1244, PMLR, 2020. arXiv:2002.03673.

[25] O. Shchur, *Modeling Continuous-time Event Data with Neural Temporal Point Processes*. PhD thesis, Technical University of Munich, Munich, Germany, 2022.

[26] Y. Dai, Y. Zhang, S. Liao, and L. Shao, "Attentional feature fusion," in *Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, (Beijing, China), pp. 1234–1244, IEEE, 2020.

[27] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *31st International Conference on Neural Information Processing Systems (NeurIPS)*, pp. 5998–6008, Curran Associates Inc., 2017.