

## 门控循环网络辨识准分子激光器能量模型

冯泽斌<sup>1,2,3</sup>, 周翊<sup>1,3\*</sup>, 江锐<sup>1,3\*\*</sup>, 韩晓泉<sup>1\*\*\*</sup>, 徐向宇<sup>1,3</sup>, 刘斌<sup>1,3</sup><sup>1</sup>中国科学院微电子研究所, 北京 100029;<sup>2</sup>中国科学院大学, 北京 100049;<sup>3</sup>北京科益虹源光电技术有限公司技术中心, 北京 100176

**摘要** 准分子激光器的放电过程是一个复杂的非线性过程, 从而导致基于放电动力学建立的激光器放电能量模型的精度很难达到仿真研究和控制算法设计的需求。通过深度学习的方法, 利用门控循环网络辨识准分子激光器放电能量模型。首先基于准分子激光器出光能量特性, 选定所建立的门控循环网络的输入。然后根据门控循环神经网络的输入特性和输出特性建立适用于准分子激光器能量模型辨识的神经网络, 并介绍了门控循环神经网络训练方法。最后利用实际采集的激光器的能量数据对门控循环神经网络进行训练。实验结果证明, 本文所设计的门控循环神经网络收敛, 辨识出来的能量模型的最大误差小于 1.5%。该方法可以应用于准分子激光器能量模型的辨识。

**关键词** 激光器; 门控循环单元网络; 准分子激光器; 模型辨识

**中图分类号** TN248.2

**文献标识码** A

**doi:** 10.3788/CJL202148.0901004

## 1 引言

随着激光技术的飞速发展, 激光技术在各个领域得到了广泛的应用<sup>[1-3]</sup>。准分子激光器以其短波长、高功率和窄线宽等特征被广泛应用于工业、医疗和科研等领域<sup>[4-6]</sup>。特别是稀有气体卤化物准分子激光器, 由于其输出激光的峰值功率高、单脉冲能量大、波长在紫外区等特点<sup>[7-8]</sup>, 成为目前半导体光刻行业最主要的激光光源<sup>[9-10]</sup>。其能量特性是光刻用准分子激光器的三大关键指标(能量、线宽、波长)之一<sup>[11]</sup>, 直接决定了半导体光刻的加工精度、产率和关键尺寸。准分子激光器的出光能量模型是研究和控制激光器能量特性的基础。

对于准分子激光器出光能量模型的研究, 国外起步比较早。Johnson 等<sup>[12]</sup>针对于大功率气体激光器提出了通用的动力学特性模型, Akashid 等<sup>[13]</sup>研究了 ArF 准分子激光器中光电离对于放电电场的影响。近些年, 国内对于准分子激光器出光能量模型的研究也逐渐增多, 文献<sup>[14]</sup>从准分子激光器放电动力学层面建立了近似的准分子激光器出光能

量模型; 文献<sup>[15]</sup>基于模糊神经网络和知识推理规则的智能控制方法对准分激光器进行控制, 得到了较好的控制效果; 文献<sup>[16]</sup>基于建立的近似出光能量模型, 设计了激光器在连续出光工作模式下的能量稳定性控制器; 文献<sup>[17]</sup>在一个经验模型上利用 PI 算法和强化学习算法相结合的方法设计了能量稳定性控制算法。从这些研究中可以看到准分子激光器出光能量模型的重要性, 而且在研究准分子激光器出光能量时, 模型越接近实际出光能量规律, 越有利于研究。

深度学习理论发展已经趋于完备, 并成为一种工具, 得到了广泛的应用<sup>[18]</sup>。其中深度循环神经网络(RNN)是深度学习领域的一个重要分支, 在语言识别、机器翻译、文本分析等领域有着广泛的应用<sup>[19]</sup>。近几年, 在国外循环神经网络, 特别是其变体——门控循环网络(GRU), 已经被应用于模型识别、趋势预测等领域。循环神经网络近些年主要应用在电池电压的建模与预测<sup>[20]</sup>、油气设备运行状态的预测<sup>[21]</sup>。GRU 由于克服了循环神经网络的长短记忆对训练的影响, 近些年被广泛应用于自然语音

收稿日期: 2020-09-14; 修回日期: 2020-10-17; 录用日期: 2020-11-23

基金项目: 国家科技重大专项(2013ZX022020)

\*E-mail: zhouyil@ime.ac.cn; \*\*E-mail: jiangrui@ime.ac.cn; \*\*\*E-mail: hanxiaoquan@ime.ac.cn

处理<sup>[22-23]</sup>等领域,同时它和其他深度学习算法相结合(如支持向量回归算法)<sup>[24]</sup>,应用于交通状况的预测、设备运行状态的监控等领域<sup>[25-26]</sup>。国内对于 GRU 的研究也逐渐成为热门,现在主要研究在语音识别、故障预测、行为预测等领域<sup>[27-28]</sup>。

本文针对准分子激光器的出光能量特性,应用 GRU 对出光能量特性进行模型辨识,通过 GRU 学习到一个准确的准分子激光器的出光能量模型。通过实验证明,学习到的模型所产生的激光脉冲能量和实际激光器产生的能量最大误差小于 1.5%。根据文献<sup>[29]</sup>中的剂量精度和能量稳定性的关系可以得出,在满足光刻用剂量精度 0.5% 时,能量稳定性的最大误差为 2.74%,该模型产生的能量的误差小于能量稳定性控制精度的误差,因此该模型满足激光器能量特性控制的仿真需求。该模型可以应用于准分子激光器出光能量特性分析、单脉冲能量控制算法设计等方面的研究。

## 2 激光器单脉冲能量特性分析

在半导体光刻应用中,准分子激光器工作在爆发(Burst)模式下。在 Burst 模式下,激光脉冲的能量序列如图 1 所示。从图中可以看出,每个 Burst 的前几个脉冲的能量会出现大小不同的超调,同时在 Burst 后面的平稳区,也会有能量的波动。在放电高压恒定的情况下,激光脉冲的能量和其在 Burst 中的位置有很大关系,所以在分析激光器能量特性时,针对 Burst 中不同位置的脉冲能量进行单独分析。

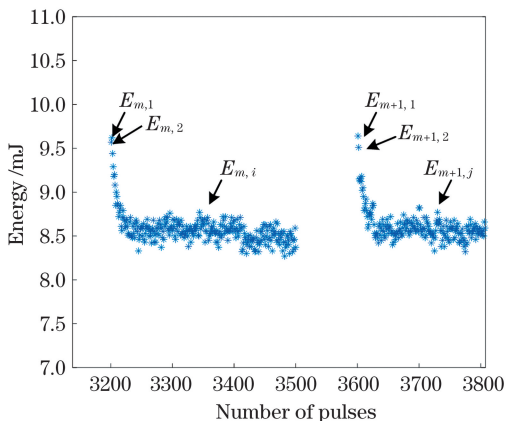


图 1 恒高压工作模式下单脉冲能量变化规律图

Fig. 1 Single pulse energy variation law under constant high voltage mode

为去除 Burst 序列中激光器脉冲能量中夹杂的噪声,分别对 Burst 中相同位置的激光脉冲能量取均值

$$\bar{E}_i = \frac{\sum_{m=1}^N E_{m,i}}{N}, \quad (1)$$

式中: $\bar{E}_i$  表示在 Burst 序列中第  $i$  个激光脉冲能量的均值; $E_{m,i}$  表示第  $m$  个 Burst 中的第  $i$  个激光脉冲的能量。

在准分子激光器的放电高压被设置为 1500 V 时,分别采集激光器工作在重复频率 1、2、3、4 kHz 工况下 1 min 的单脉冲能量数据。对于不同工作重复频率的激光单脉冲能量数据,分别对 Burst 中相同位置的激光脉冲能量取均值,得出 Burst 中激光脉冲能量的变化规律如图 2 所示。准分子激光的能量变化趋势及其他统计特性见文献<sup>[29]</sup>。

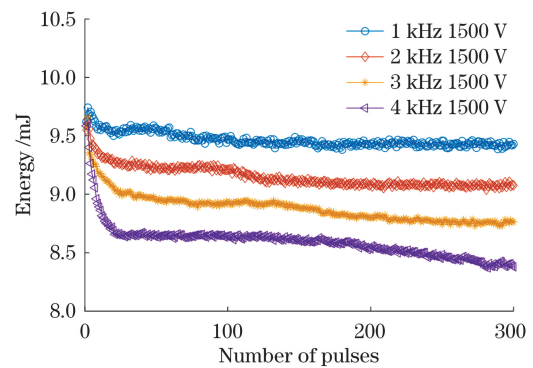


图 2 恒高压模式下不同重复频率 Burst 脉冲能量变化规律

Fig. 2 Energy variation of pulses in Burst with different repetition rates under constant high voltage mode

## 3 门控循环网络辨识算法建立

### 3.1 循环神经网络结构

循环神经网络表示的是一系列时刻变化的模型,循环神经网络的结构如图 3 所示。图中左侧为循环神经网络的统一描述,表示在每个时刻的网络模型,将其展开可以表示成图 3 中右侧的结构,图中每个结点代表若干神经结点。其中  $\mathbf{x}(t)$  表示在  $t$  时刻的输入,  $\mathbf{W}$  表示输入层权重矩阵,  $\mathbf{S}(t)$  表示循环神经网络隐藏层的状态,  $\mathbf{U}$  表示神经网络隐藏层间传递的权重矩阵,  $\mathbf{Y}(t)$  表示  $t$  时刻网络的输出,  $\mathbf{V}$  表示输出层的权重矩阵,  $\mathbf{E}(t)$  表示样本在  $t$  时刻的值,  $L(t)$  表示  $t$  时刻衡量网络输出准确性损失函数。通过循环网络结构可以看出,网络在某一时刻的输入不仅与输入有关,还与历史时刻隐藏层的状态有关,网络中的权重矩阵参数在整个网络中可以共享。循环神经网络的结构特性决定了其特别适用于处理序惯性问题。

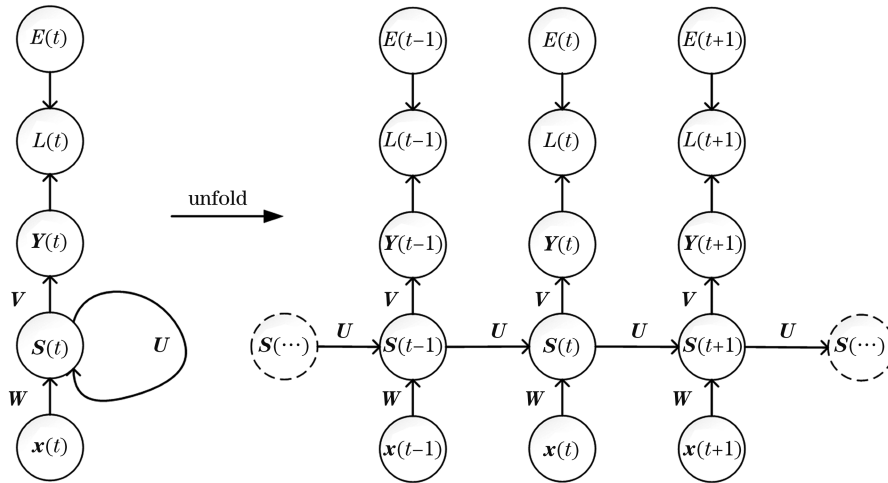


图 3 循环神经网络结构图

Fig. 3 Recurrent neural network structure diagram

### 3.2 门控循环网络结构

循环神经网络的序列中层数较多时,在训练过程中由于长期依赖问题的存在,会出现梯度爆炸或梯度消失的现象,从而导致训练的结果不收敛。为解决该问题,长短期记忆网络(LSTM)在1997年被首次提出<sup>[30]</sup>。LSTM是一种时间循环神经网络,是为了解决一般的循环神经网络存在的长期依赖问题而专门设计出来的,但是LSTM计算复杂。门控循环网络也是一种循环神经网络,和LSTM一样,也是为了解决长期记忆和反向传播中的梯度消失等问题而提出来的<sup>[24]</sup>。GRU与LSTM相比,在很大程度上减少了计算量,更适用于解决实际工程问题。

GRU的每个序列隐藏层的结构如图4所示。图中 $h(t-1)$ 表示前一时间步隐藏层的状态, $h(t)$ 表示当前时间步隐藏层的状态, $x(t)$ 表示当前时间步的输入。从图中可以看出,GRU的隐藏层在利用上一个隐藏层状态的时候,对其进行了处理。主要是增加了两个门限,一个是重置门 $r(t)$ ,另一个是更新门

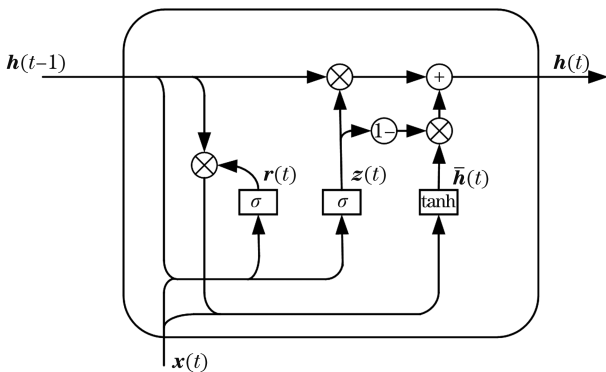


图 4 隐藏层网络结构图

Fig. 4 Hidden layer network structure diagram

$z(t)$ 。重置门表示当前状态忽略前一时间步状态的程度,用于捕捉时间序列里短期的依赖关系,更新门表示前一时间步的状态代入这一时间步的信息量,用于捕捉时间序列里长期的依赖关系。 $\bar{h}(t)$ 为候选隐藏层状态,用于辅助隐藏层状态 $h(t)$ 的计算。

### 3.3 辨识激光器出光能量模型的GRU神经网络设计

为了利用GRU辨识激光器的出光能量模型,需要设计出适用于实际应用的GRU。通过对准分子激光器出光能量特性的分析可知,直接影响激光器出光能量的条件变量为激光脉冲的时间间隔和放电高压值。所以GRU的输入变量选定为激光脉冲的时间间隔和放电高压值,表示为

$$x(t) = W_{in} [V_H(t) \Delta t(t)], \quad (2)$$

式中: $x(t)$ 表示 $t$ 时刻的GRU输入; $V_H(t)$ 表示 $t$ 时刻放电高压值; $\Delta t(t)$ 表示 $t$ 时刻的激光脉冲的时间间隔; $W_{in}$ 表示输入尺度变换矩阵。

更新门的激活函数为sigmoid函数,表示为 $\sigma$ ,则更新门的表达式为

$$z(t) = \sigma[W_z \cdot x(t) + U_z \cdot h(t-1)], \quad (3)$$

式中: $W_z$ 表示更新门的输入权重矩阵; $U_z$ 表示更新门的隐藏层状态传递矩阵。

重置门的表达式为

$$r(t) = \sigma[W_r \cdot x(t) + U_r \cdot h(t-1)], \quad (4)$$

式中: $W_r$ 表示重置门的输入权重矩阵; $U_r$ 表示重置门的隐藏层状态传递矩阵。

候选隐藏层状态的激活函数选用tanh函数,其表达式为

$$\bar{h}(t) = \tanh\{W \cdot x(t) + U \cdot [r(t) \odot h(t-1)]\}, \quad (5)$$

式中:  $\mathbf{W}$  表示候选隐藏层状态的输入权重矩阵;  $\mathbf{U}$  表示候选隐藏层状态针对上一时刻隐藏层状态的传递矩阵;  $\odot$  表示哈达玛积。

当前时刻的隐藏层状态可以表示为

$$\mathbf{h}(t) = [1 - \mathbf{z}(t)] \odot \mathbf{h}(t-1) + \mathbf{z}(t) \odot \bar{\mathbf{h}}(t). \quad (6)$$

输出层的激活函数选用 sigmoid 函数, 则输出层可以表示为

$$y(t) = \sigma[\mathbf{W}_y \cdot \mathbf{h}(t)], \quad (7)$$

式中:  $\mathbf{W}_y$  表示隐藏层状态到输出层的权重矩阵。

通过网络得出的每个脉冲的能量可以表示为

$$E(t) = \mathbf{W}_E \cdot y(t), \quad (8)$$

式中:  $\mathbf{W}_E$  表示输出尺度变换系数。

综合(2)~(8)式, 构建了一个用于激光器出光能量模型识别的 GRU 网络模型, 即通过数据训练确定其中的参数后, (2)~(8)式就构成了一个完整的准分子激光器出光能量模型。

### 3.4 门控循环网络的训练方法

因为建立的 GRU 网络中的各个权重矩阵需要通过学习获得, 所以需要 GRU 网络进行训练。在实际应用中, 采用的训练方法为时间反向传播 (BPTT) 算法。

GRU 网络应用到激光器出光能量模型辨识中,

以一个 Burst 作为一个时间序列, 定义整个网络的损失函数为

$$l = \sum_{t=1}^n \frac{1}{2} [E_t(t) - E(t)]^2, \quad (9)$$

式中:  $E_t(t)$  为  $t$  时刻的训练样本的激光单脉冲能量值。

令  $y_{\text{net}}(t) = \mathbf{W}_y \cdot \mathbf{h}(t)$ , 则

$$\frac{\partial l}{\partial y_{\text{net}}(t)} = \frac{\partial l}{\partial E(t)} \cdot \frac{\partial E(t)}{\partial y(t)} \cdot \frac{\partial y(t)}{\partial y_{\text{net}}(t)} = [E_t(t) - E(t)] \cdot \mathbf{W}_E \cdot \sigma'[y_{\text{net}}(t)]. \quad (10)$$

那么可以得到

$$\frac{\partial l}{\partial \mathbf{W}_y} = \sum_{t=1}^n \left\{ \frac{\partial l}{\partial y_{\text{net}}(t)} \cdot \left[ \frac{\partial y_{\text{net}}(t)}{\partial \mathbf{W}_y} \right]^T \right\} = \sum_{t=1}^n \left\{ [E_t(t) - E(t)] \cdot \mathbf{W}_E \cdot \sigma'[y_{\text{net}}(t)] \cdot \mathbf{h}^T(t) \right\}, \quad (11)$$

式中:  $\sigma'[y_{\text{net}}(t)]$  表示  $\sigma[y_{\text{net}}(t)]$  的导数。

当  $t=n$  时, 有

$$\frac{\partial l}{\partial \mathbf{h}(t)} = \left[ \frac{\partial y_{\text{net}}(t)}{\partial \mathbf{h}(t)} \right]^T \cdot \frac{\partial l}{\partial y_{\text{net}}(t)} = \mathbf{W}_y^T \cdot [E_t(t) - E(t)] \cdot \mathbf{W}_E \cdot \sigma'[y_{\text{net}}(t)]. \quad (12)$$

当  $t < n$  时, 有

$$\frac{\partial l}{\partial \mathbf{h}(t)} = \left[ \frac{\partial y_{\text{net}}(t)}{\partial \mathbf{h}(t)} \right]^T \cdot \frac{\partial l}{\partial y_{\text{net}}(t)} + \left[ \frac{\partial \mathbf{z}(t+1)}{\partial \mathbf{h}(t)} \right]^T \cdot \frac{\partial l}{\partial \mathbf{z}(t+1)} + \left[ \frac{\partial \mathbf{r}(t+1)}{\partial \mathbf{h}(t)} \right]^T \cdot \frac{\partial l}{\partial \mathbf{r}(t+1)} + \left[ \frac{\partial \bar{\mathbf{h}}(t+1)}{\partial \mathbf{h}(t)} \right]^T \cdot \frac{\partial l}{\partial \bar{\mathbf{h}}(t+1)} + \left[ \frac{\partial \mathbf{h}(t+1)}{\partial \mathbf{h}(t)} \right]^T \cdot \frac{\partial l}{\partial \mathbf{h}(t+1)}, \quad (13)$$

其中

$$\frac{\partial l}{\partial \mathbf{z}(t+1)} = \frac{\partial l}{\partial \mathbf{h}(t+1)} \odot \frac{\partial \mathbf{h}(t+1)}{\partial \mathbf{z}(t+1)} = \frac{\partial l}{\partial \mathbf{h}(t+1)} \odot [\bar{\mathbf{h}}(t+1) - \mathbf{h}(t)], \quad (14)$$

$$\frac{\partial l}{\partial \bar{\mathbf{h}}(t+1)} = \frac{\partial l}{\partial \mathbf{h}(t+1)} \odot \frac{\partial \mathbf{h}(t+1)}{\partial \bar{\mathbf{h}}(t+1)} = \frac{\partial l}{\partial \mathbf{h}(t+1)} \odot \mathbf{z}(t+1), \quad (15)$$

$$\frac{\partial l}{\partial \mathbf{r}(t+1)} = \left[ \frac{\partial \bar{\mathbf{h}}(t+1)}{\partial \mathbf{r}(t+1)} \right]^T \cdot \frac{\partial l}{\partial \bar{\mathbf{h}}(t+1)} = \text{diag}\{\tanh'[\bar{\mathbf{h}}_{i\text{Net}}(t+1)]\} \cdot \mathbf{U}^T \odot \mathbf{h}(t) \cdot \frac{\partial l}{\partial \mathbf{h}(t+1)}. \quad (16)$$

通过(12)~(16)式可以看出, 当  $t < n$  时,  $\frac{\partial l}{\partial \mathbf{h}(t)}$  是一个有关  $\frac{\partial l}{\partial \mathbf{h}(t+1)}$  的递推表达式, 所以在训练时, 从  $t=n$  开始反向计算  $\frac{\partial l}{\partial \mathbf{h}(t)}$ , 与此同时, 可

以计算出  $\frac{\partial l}{\partial \mathbf{z}(t)}$ 、 $\frac{\partial l}{\partial \bar{\mathbf{h}}(t)}$ 、 $\frac{\partial l}{\partial \mathbf{r}(t)}$ 。

令  $\mathbf{z}_{\text{net}}(t) = \mathbf{W}_z \cdot \mathbf{x}(t) + \mathbf{U}_z \cdot \mathbf{h}(t-1)$ , 则针对  $\mathbf{W}_z$  的梯度, 有

$$\frac{\partial l}{\partial \mathbf{W}_z} = \sum_{t=0}^n \left\{ \left[ \frac{\partial \mathbf{z}(t)}{\partial \mathbf{W}_z} \right]^T \cdot \frac{\partial l}{\partial \mathbf{z}(t)} \right\} = \sum_{t=0}^n \left\{ \text{diag}\{\sigma'[\mathbf{z}_{i\text{Net}}(t)]\} \cdot \frac{\partial l}{\partial \mathbf{z}(t)} \cdot \mathbf{x}^T(t) \right\}. \quad (17)$$

针对  $\mathbf{U}_z$  的梯度, 有

$$\frac{\partial l}{\partial \mathbf{U}_z} = \sum_{t=0}^n \left\{ \left[ \frac{\partial \mathbf{z}(t)}{\partial \mathbf{U}_z} \right]^T \cdot \frac{\partial l}{\partial \mathbf{z}(t)} \right\} = \sum_{t=0}^n \left\{ \text{diag}\{\sigma'[\mathbf{z}_{i\text{Net}}(t)]\} \cdot \frac{\partial l}{\partial \mathbf{z}(t)} \cdot \mathbf{h}^T(t-1) \right\}, \quad (18)$$

式中:  $\text{diag}(\cdot)$  表示对角矩阵;  $\mathbf{z}_{i\text{Net}}(t)$  表示  $\mathbf{z}_{\text{net}}(t)$  的第  $i$  个分量。

令  $\bar{\mathbf{h}}_{\text{net}}(t) = \mathbf{W} \cdot \mathbf{x}(t) + \mathbf{U} \cdot [\mathbf{r}(t) \odot \mathbf{h}(t-1)]$ , 则针对  $\mathbf{W}$  的梯度, 有

$$\frac{\partial l}{\partial \mathbf{W}} = \sum_{t=0}^n \left\{ \left[ \frac{\partial \bar{\mathbf{h}}(t)}{\partial \mathbf{W}} \right]^T \cdot \frac{\partial l}{\partial \bar{\mathbf{h}}(t)} \right\} = \sum_{t=0}^n \left\{ \text{diag}\{\tanh'[\bar{\mathbf{h}}_{i\text{Net}}(t)]\} \cdot \frac{\partial l}{\partial \bar{\mathbf{h}}(t)} \cdot \mathbf{x}^T(t) \right\}. \quad (19)$$

针对  $\mathbf{U}$  的梯度, 有

$$\frac{\partial l}{\partial \mathbf{U}} = \sum_{t=0}^n \left\{ \left[ \frac{\partial \bar{\mathbf{h}}(t)}{\partial \mathbf{U}} \right]^T \cdot \frac{\partial l}{\partial \bar{\mathbf{h}}(t)} \right\} = \sum_{t=0}^n \left\{ \text{diag}\{\tanh'[\bar{\mathbf{h}}_{i\text{Net}}(t)]\} \cdot \frac{\partial l}{\partial \bar{\mathbf{h}}(t)} \cdot [\mathbf{r}(t) \odot \mathbf{h}(t-1)]^T \right\}, \quad (20)$$

式中:  $\tanh'(\cdot)$  表示  $\tanh(\cdot)$  的导数;  $\bar{\mathbf{h}}_{i\text{Net}}(t)$  表示  $\bar{\mathbf{h}}_{\text{net}}(t)$  的第  $i$  个分量。

令  $\mathbf{r}_{\text{net}}(t) = \mathbf{W}_r \cdot \mathbf{x}(t) + \mathbf{U}_r \cdot \mathbf{h}(t-1)$ , 则针对  $\mathbf{W}_r$  的梯度, 有

$$\frac{\partial l}{\partial \mathbf{W}_r} = \sum_{t=0}^n \left\{ \left[ \frac{\partial \mathbf{r}(t)}{\partial \mathbf{W}_r} \right]^T \cdot \frac{\partial l}{\partial \mathbf{r}(t)} \right\} = \sum_{t=0}^n \left\{ \text{diag}\{\sigma'[\mathbf{r}_{i\text{Net}}(t)]\} \cdot \frac{\partial l}{\partial \mathbf{r}(t)} \cdot \mathbf{x}^T(t) \right\}. \quad (21)$$

针对  $\mathbf{U}_r$  的梯度, 有

$$\frac{\partial l}{\partial \mathbf{U}_r} = \sum_{t=0}^n \left\{ \left[ \frac{\partial \mathbf{r}(t)}{\partial \mathbf{U}_r} \right]^T \cdot \frac{\partial l}{\partial \mathbf{r}(t)} \right\} = \sum_{t=0}^n \left\{ \text{diag}\{\sigma'[\mathbf{r}_{i\text{Net}}(t)]\} \cdot \frac{\partial l}{\partial \mathbf{r}(t)} \cdot \mathbf{h}^T(t-1) \right\}, \quad (22)$$

式中:  $\mathbf{r}_{i\text{Net}}(t)$  表示  $\mathbf{r}_{\text{net}}(t)$  的第  $i$  个分量。

由(9)~(22)式可以得到 GRU 神经网络的训练参数的更新算法,

$$\mathbf{W}_y(i+1) = \mathbf{W}_y(i) - \lambda \cdot \frac{\partial l}{\partial \mathbf{W}_y}, \quad (23)$$

$$\mathbf{W}_z(i+1) = \mathbf{W}_z(i) - \lambda \cdot \frac{\partial l}{\partial \mathbf{W}_z}, \quad (24)$$

$$\mathbf{U}_z(i+1) = \mathbf{U}_z(i) - \lambda \cdot \frac{\partial l}{\partial \mathbf{U}_z}, \quad (25)$$

$$\mathbf{W}(i+1) = \mathbf{W}(i) - \lambda \cdot \frac{\partial l}{\partial \mathbf{W}}, \quad (26)$$

$$\mathbf{U}(i+1) = \mathbf{U}(i) - \lambda \cdot \frac{\partial l}{\partial \mathbf{U}}, \quad (27)$$

$$\mathbf{W}_r(i+1) = \mathbf{W}_r(i) - \lambda \cdot \frac{\partial l}{\partial \mathbf{W}_r}, \quad (28)$$

$$\mathbf{U}_r(i+1) = \mathbf{U}_r(i) - \lambda \cdot \frac{\partial l}{\partial \mathbf{U}_r}. \quad (29)$$

#### 4 准分子激光器出光能量模型学习

利用 GRU 学习准分子激光器的出光能量模型需要大量的数据。数据取自一台产生 248 nm 的 KrF 准分子激光器, 其工作在 4 kHz 的重复频率。由于准分子激光的波长也会对能量数据有影响, 所以在实验过程中, 将波长利用反馈技术控制在 248.327 nm。分别在 1400、1450、1550、1600 V 的放电高压下, 采集激光器的单脉冲激光能量 1 min。对数据按(1)式进行预处理。这样使得训练 GRU 网络所采用的能量数据是在每一个放电高压下激光

器 Burst 中的相同位置激光脉冲能量的均值。为了充分利用数据, 在每一次训练时, 随机选取不同放电高压下的数据。设定终止条件为 10 万次训练或 GRU 构建的模型的能量最大偏差小于 0.15 mJ。GRU 学习准分子激光器出光模型的程序流程图如图 5 所示。

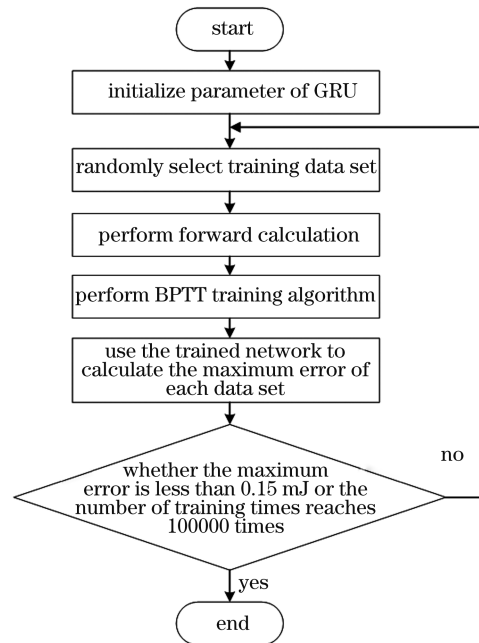


图 5 利用 GRU 网络学习激光器出光能量模型程序流程图  
Fig. 5 Program flow chart of learning laser energy model using GRU network

通过训练可以得到最终模型的参数  $\mathbf{W}_y$ 、 $\mathbf{W}_z$ 、 $\mathbf{U}_z$ 、 $\mathbf{W}$ 、 $\mathbf{U}$ 、 $\mathbf{W}_r$ 、 $\mathbf{U}_r$ , 从而(2)~(8)式构成了准分子激

激光器出光能量模型。得到的 GRU 神经网络计算出的各个电压下的能量与实际数据集能量间误差如图 6 所示。每个高压下 Burst 中激光脉冲的能量误差均小于 0.15 mJ, 由于能量中心值为 10 mJ, 所以相对误差小于 1.5%。

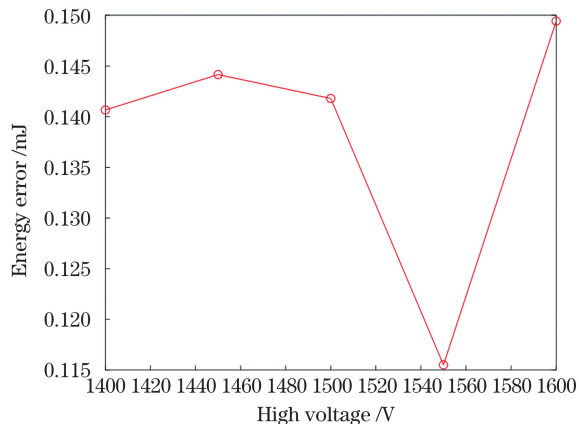


图 6 各电压下 GRU 计算得到的能量和实际数据集能量间的最大误差

Fig. 6 Maximum error between the energy calculated by GRU and the energy of the actual data set at each voltage

训练过程中, 能量最大误差的变化如图 7 所示, 横坐标的步长为神经网络训练的次数, 纵坐标为能量误差的绝对值。从图中可以看出, 最大误差在逐渐减小, 直到小于 0.15 mJ, 证明本研究构建的 GRU 神经网络在训练过程中是收敛的。

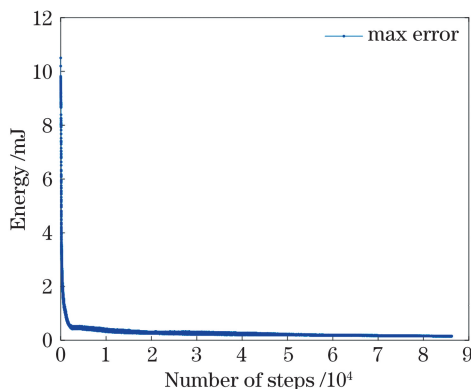


图 7 训练过程中最大能量误差变化曲线

Fig. 7 Maximum energy error curve during training

选用训练集之外的数据对模型进行验证。用训练得到的模型对高压为 1550 V 时的激光器出光能量进行计算, 得出的能量值和实际激光器上采集到的能量值经过(1)式处理后的对比如图 8 所示。图 8(a)表示 GRU 网络得到的能量变化与实际数据的能量变化的重合度, 图 8(b)表示实际数据的能量变化, 图 8(c)表示 GRU 网络得到的能量变化。从图

中可以看出, 通过 GRU 神经网络得到的能量变化与实际脉冲的能量变化有较好的重合度。

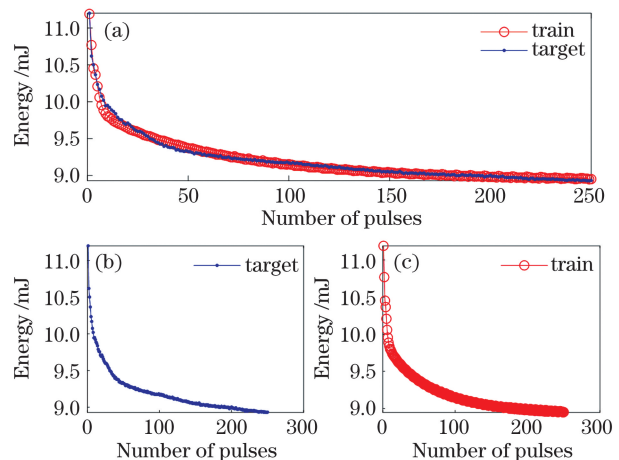


图 8 电压为 1550 V 时 GRU 网络得到的单脉冲能量与实际数据集单脉冲能量对比图。(a)测试数据与实际数据重合图;(b)实际数据图;(c)测试数据图

Fig. 8 Comparison graph of single pulse energy obtained by GRU network and single pulse energy of actual data set when voltage is 1550 V. (a) Coincidence graph of test data and actual data; (b) actual data graph; (c) test data graph

激光器工作在恒高压 1500 V 工况下, 分别取工作重复频率在 1、2、3、4 kHz 的激光单脉冲能量数据。利用(1)式对于数据进行预处理, 得到的数据如图 9(a)所示。训练得到的模型产生的数据如图 9(b)所示。对比图 9(a)和(b)可以看到, 模型得到的数据在不同重复频率下, Burst 中激光能量变化与实际测得的激光器能量数据趋势一致, 其最大误差如图 10 所示, 在图中可以看到, 在不同重复频率下最大误差小于 0.13 mJ, 即小于 1.5%。

## 5 结 论

准分子激光器的出光能量模型是一个复杂的非线性模型, 很难从理论推导得出一个精确的模型。文献[14]虽然从放电动力学角度建立了一个准分子激光器的能量理论模型, 但是该模型只能进行理论分析。文献[29]中虽然建立了一个准分子激光器的经验模型, 能够进行控制算法的仿真研究, 但是在建模过程中做了多个假设, 该模型产生的能量序列和实际的准分子激光器出光能量序列还有很大差距。然而实际的研究和应用工作需要一个准确的激光器出光能量模型。本文通过深度学习的方法, 利用 GRU 神经网络, 对激光波长处于 248.327 nm, 放电重复频率为 4 kHz 的激光器在不同放电高压下采

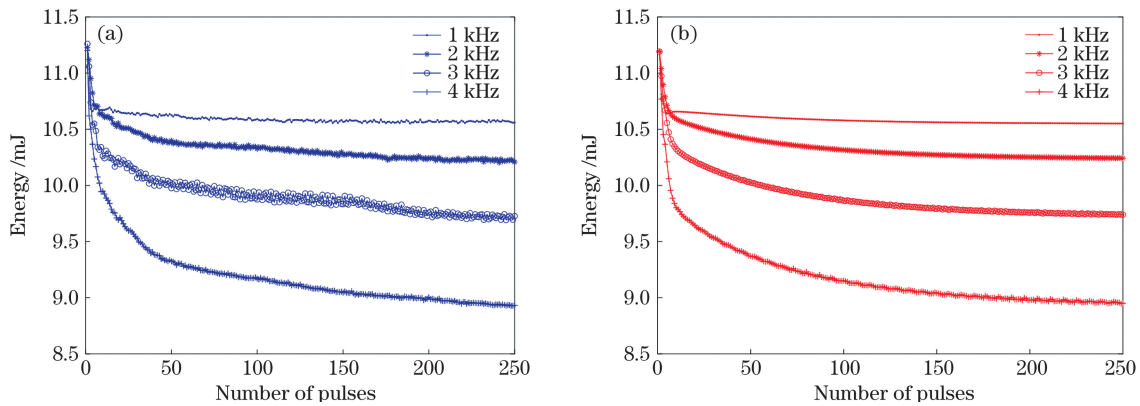


图 9 恒高压 1500 V 工况下, 1 kHz 到 4 kHz 重复频率的激光器能量和模型产生能量对比图。(a) 测试数据图; (b) 实际数据图

Fig. 9 Comparison of laser energy and model generated energy between 1 kHz to 4 kHz repetition frequency under constant high voltage 1500 V. (a) Test data graph; (b) actual data graph

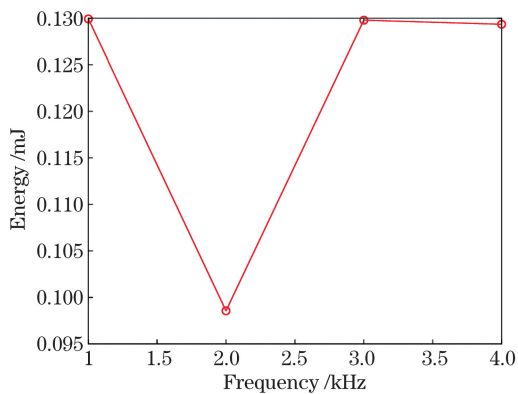


图 10 恒高压 1500 V 工况下, 1 kHz 到 4 kHz 重复频率的激光器能量和模型产生能量的最大误差

Fig. 10 Maximum error of laser energy and model generated energy between 1 kHz to 4 kHz repetition frequency under constant high voltage 1500 V

集的激光能量数据集进行学习, 辨识到能量模型。然后在激光波长处于 248.327 nm, 分别利用不同重复频率和不同放电高压下的激光器出光能量数据集进行验证。验证结果显示, 通过 GRU 神经网络辨识出来的激光器出光能量模型产生的脉冲能量与实际能量的最大误差小于 1.5%。该精度小于剂量控制中要求能量的稳定性 2.74%, 满足模型的控制效果仿真需求。该方法能够比较精确地辨识激光器出光能量模型, 利用辨识的模型能够比较方便地对能量控制算法进行仿真研究, 方便算法分析, 缩短了实验时间, 从而对提高准分子激光器的能量稳定性控制和剂量精度的控制有很大的意义。

### 参 考 文 献

[1] Zhang X C, Qian J, Fu Q, et al. Cutting of PBO

fiber-reinforced composites using picosecond lasers [J]. Chinese Journal of Lasers, 2020, 47(10): 1002004.

张学聪, 钱静, 付强, 等. 皮秒激光切割 PBO 纤维增强复合材料 [J]. 中国激光, 2020, 47(10): 1002004.

[2] Qi N J, Yuan X D, Zhang L J, et al. Three-dimensional detection technology of laser damage residual stress [J]. Chinese Journal of Lasers, 2020, 47(10): 1004001.

齐乃杰, 袁晓东, 张丽娟, 等. 激光损伤残余应力三维检测技术 [J]. 中国激光, 2020, 47(10): 1004001.

[3] Zhang Y X, Yu H H, Zhang H J, et al. Research progress of ultrafast pulsed lasers in visible range [J]. Chinese Journal of Lasers, 2019, 46(5): 0508011.

张玉霞, 于浩海, 张怀金, 等. 可见波段超快脉冲激光研究进展 [J]. 中国激光, 2019, 46(5): 0508011.

[4] Pan N, Liang X, Lin Y, et al. Transmission method of analog signal in excimer laser system [J]. Infrared and Laser Engineering, 2019, 48(9): 0905003.

潘宁, 梁勛, 林颖, 等. 准分子激光系统中模拟信号的传输方法 [J]. 红外与激光工程, 2019, 48(9): 0905003.

[5] Zhu F, Yu L, Yi A P, et al. Spectral compression of XeF(C-A) excimer laser using grating resonator with an intra-cavity etalon [J]. Optics and Precision Engineering, 2018, 26(12): 2888-2893.

朱峰, 于力, 易爱平, 等. 内置标准具光栅腔 XeF(C-A) 准分子激光器的线宽压缩 [J]. 光学精密工程, 2018, 26(12): 2888-2893.

[6] Zhou C Q, Yu L, Lu P H, et al. Effect of asphericity parameter on Seidal aberration and ablation depth in laser refractive surgery [J]. Optics and Precision Engineering, 2007, 15(2): 167-172.

周传清, 余雷, 陆培华, 等. 准分子屈光手术中非球面系数对球差以及切削深度的影响 [J]. 光学精密工

- 程, 2007, 15(2): 167-172.
- [7] Bi D D, Zhang L C, Shi G, et al. Optical coatings for projection objective immersion lithography [J]. *Chinese Optics*, 2018, 11(5): 745-764.  
毕丹丹, 张立超, 时光, 等. 浸没式光刻投影物镜光学薄膜[J]. *中国光学*, 2018, 11(5): 745-764.
- [8] Petrova T B, Petrov G M, Wolford M F, et al. Modeling of an electron-beam pumped ArF excimer laser [C] // 2017 IEEE International Conference on Plasma Science (ICOPS), May 21-25, 2017, Atlantic City, NJ, USA. New York: IEEE Press, 2017: 1.
- [9] Tsushima H, Fujimaki Y, Kiyota Y, et al. Extremely long life excimer laser chamber technology for multi-patterning lithography [J]. *Proceedings of SPIE*, 2019, 10961: 109610X.
- [10] Losev V F, Panchenko Y N, Dudarev V V, et al. Development of discharge pulse repetition rate excimer lasers for different applications [J]. *Chinese Journal of Optics*, 2011, 4(1): 1-8.
- [11] Yuan J C, Zhao J S, Li H, et al. Research of peak-detection algorithm based on absolute wavelength calibration of excimer laser [J]. *Chinese Journal of Lasers*, 2018, 45(7): 0701003.  
袁靖超, 赵江山, 李慧, 等. 基于准分子激光绝对波长校准的寻峰算法研究 [J]. *中国激光*, 2018, 45(7): 0701003.
- [12] Johnson T, Palumbo L, Hunter A, et al. Kinetics simulation of high-power gas lasers [J]. *IEEE Journal of Quantum Electronics*, 1979, 15(5): 289-301.
- [13] Akashi H, Sakai Y, Tagashira I, et al. Modelling of a self-sustained discharge-excited ArF excimer laser: the influence of photo-ionization and photodetachment by laser light on the discharge development [J]. *Journal of Physics D: Applied Physics*, 1995, 28(3): 445-451.
- [14] Wang Q, Zhao J S, Luo S W, et al. Energy efficiency analysis of ArF excimer laser system [J]. *Acta Physica Sinica*, 2016, 65(21): 214205.  
王倩, 赵江山, 罗时文, 等. ArF 准分子激光系统的能量效率特性 [J]. *物理学报*, 2016, 65(21): 214205.
- [15] Fan Y P, Huang X Y, Yu Y S, et al. Adjusting-controlling on gassing and energy for high power excimer laser based on intelligence [J]. *Control Theory & Applications*, 2002, 19(4): 561-566.  
樊友平, 黄席樾, 余吟山, 等. 基于智能的大功率准分子激光器配气与能量的调控 [J]. *控制理论与应用*, 2002, 19(4): 561-566.
- [16] Wang X S, Liang X, You L B, et al. Study on energy control algorithm for high-repetition-rate ArF excimer lasers [J]. *Laser Technology*, 2012, 36(6): 763-766.  
王效顺, 梁勛, 游利兵, 等. 高重复频率 ArF 准分子激光器能量控制算法研究 [J]. *激光技术*, 2012, 36(6): 763-766.
- [17] Sun Z X, Feng Z B, Zhou Y, et al. Energy control of excimer laser based on reinforcement learning [J]. *Chinese Journal of Lasers*, 2020, 47(9): 0901002.  
孙泽旭, 冯泽斌, 周翊, 等. 基于强化学习的准分子激光器能量控制算法研究 [J]. *中国激光*, 2020, 47(9): 0901002.
- [18] Sun Z J, Xue L, Xu Y M, et al. Overview of deep learning [J]. *Application Research of Computers*, 2012, 29(8): 2806-2810.  
孙志军, 薛磊, 许阳明, 等. 深度学习研究综述 [J]. *计算机应用研究*, 2012, 29(8): 2806-2810.
- [19] Zhang J, Qu D, Li Z, et al. Recurrent neural network language model based on word vector features [J]. *Pattern Recognition and Artificial Intelligence*, 2015, 28(4): 299-305.  
张剑, 屈丹, 李真, 等. 基于词向量特征的循环神经网络语言模型 [J]. *模式识别与人工智能*, 2015, 28(4): 299-305.
- [20] Chmielewski A, Możaryn J, Piórkowski P, et al. Battery voltage estimation using NARX recurrent neural network model [M] // Szweczyk R, Zieliński C, Kaliczyńska M, et al. *Automation 2019. Advances in intelligent systems and computing*. Cham: Springer, 2020, 920: 218-231.
- [21] Abbasi T, Lim K H, Yam K S, et al. Predictive maintenance of oil and gas equipment using recurrent neural network [J]. *IOP Conference Series: Materials Science and Engineering*, 2019, 495: 012067.
- [22] Rana R. Gated recurrent unit (GRU) for emotion classification from noisy speech [EB/OL]. (2016-12-13) [2020-09-14]. <https://arxiv.org/abs/1612.07778v1>.
- [23] Chung J, Gulcehre C, Cho K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling [EB/OL]. (2014-12-11) [2020-09-14]. <https://arxiv.org/abs/1412.3555>.
- [24] Agarap A F M. A neural network architecture combining gated recurrent unit (GRU) and support vector machine (SVM) for intrusion detection in network traffic data [C] // *Proceedings of the 2018 10th International Conference on Machine Learning and Computing*, February 26, 2018, Macao, China. New York: ACM, 2018: 26-30.
- [25] Zhao R, Wang D Z, Yan R Q, et al. Machine health monitoring using local feature-based gated recurrent unit networks [J]. *IEEE Transactions on Industrial Electronics*, 2018, 65(2): 1539-1548.



- [26] Amihai I, Chioua M, Gitzel R, et al. Modeling machine health using gated recurrent units with entity embeddings and K-means clustering [C]//2018 IEEE 16th International Conference on Industrial Informatics (INDIN), July 18-20, 2018, Porto. New York: IEEE Press, 2018: 212-217.
- [27] Niu Z W, Yu Z Y, Li B, et al. Short-term wind power forecasting model based on deep gated recurrent unit neural network [J]. Electric Power Automation Equipment, 2018, 38(5): 36-42.  
牛哲文, 余泽远, 李波, 等. 基于深度门控循环单元神经网络的短期风功率预测模型 [J]. 电力自动化设备, 2018, 38(5): 36-42.
- [28] Zhang Z H, Sun S F. Research on speech recognition based on gated recurrent unit [J]. Information & Communications, 2019, 32(3): 126-128.  
张忠豪, 孙水发. 基于门控循环单元的语音识别研究 [J]. 信息通信, 2019, 32(3): 126-128.
- [29] Feng Z B, Zhou Y, Jiang R, et al. Energy control algorithm of high frequency excimer laser [J]. Infrared and Laser Engineering, 2020, 49(11): 164-175.  
冯泽斌, 周翊, 江锐, 等. 高重复频率 KrF 准分子激光器能量特性控制 [J]. 红外与激光工程, 2020, 49(11): 164-175.
- [30] Hochreiter S, Schmidhuber J. LSTM can solve hard long time lag problems [C]//NIPS'96: Proceedings of the 9th International Conference on Neural Information Processing Systems, December 3, 1996, Cambridge, MA, United States. Cambridge: MIT Press, 1997: 473-479.

## Recognition of Energy Model of Excimer Laser by Gate Recurrent Unit

Feng Zebin<sup>1,2,3</sup>, Zhou Yi<sup>1,3\*</sup>, Jiang Rui<sup>1,3\*\*</sup>, Han XiaoQuan<sup>1,3\*\*\*</sup>, Xu Xiangyu<sup>1,3</sup>,  
Liu Bin<sup>1,3</sup>

<sup>1</sup> Institute of Microelectronics of Chinese Academy of Sciences, Beijing 100029, China;

<sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, China;

<sup>3</sup> Beijing RSLaser Opto-Electronics Technology Co., Ltd., Beijing 100176, China

### Abstract

**Objective** Excimer lasers are widely used in industrial, medical, and scientific fields because of their short wavelength, high power, and narrow line width. Especially rare gas halogen excimer laser, because of its high peak output power, high single pulse energy, and ultraviolet wavelength, has become the main laser source in the semiconductor lithography industry. Its energy is one of the three key parameters (energy, linewidth, and wavelength) of excimer laser for photolithography, which directly determines the processing accuracy, yield, and key dimensions of semiconductor lithography. When studying the energy of an excimer laser, the closer the model approaches the actual law of light output energy, the more conducive to the study. The output energy model of an excimer laser is the basis for studying and controlling the energy characteristics of the laser. Discharge process of excimer laser is a complex nonlinear process, which leads to the accuracy of laser discharge energy model based on discharge dynamics is difficult to meet the needs of simulation research and control algorithm design. In this paper, the method based on deep learning was applied to identify the energy mode of excimer laser to avoid the inaccuracy of theoretical modeling.

**Methods** The development of deep learning theory has become more and more complete. It has become a tool and has been widely applied. Among them, recurrent neural network (RNN) is an important branch in the field of deep learning. It has been widely used in language recognition, machine translation, text analysis and other fields. In recent years, circulating neural networks abroad, especially its variant gate recurrent unit (GRU), has been applied to model recognition, trend prediction and other fields. In this paper, the gated recurrent unit network was used to identify the discharge energy model of the excimer laser. Firstly, based on the characteristics of the excimer laser energy, the discharge voltage and discharge interval were selected as the input of the established gating recurrent unit network. Then, according to the characteristics of the gated recurrent unit network and the excimer laser energy, a neural network suitable for energy model identification of excimer laser was established. When using the GRU network to identify the laser light energy model, a burst pulse energy sequence was used as a time sequence. Finally, the back propagation through time (BPTT) was used to train the established GRU network.

**Results and Discussions** Using GRU to learn the energy model of excimer laser requires a lot of data. The data

was taken from a KrF excimer laser that produces laser of 248 nm, which worked at a repetition frequency of 4 kHz. Since the wavelength of the excimer laser also affects the energy data, in the course of the experiment, the wavelength was controlled at 248.327 nm using feedback technology. Energy data of the laser was collected under discharge high voltages of 1400 V, 1450 V, 1550 V, and 1600 V, respectively. In order to make full use of the data, at each training, the data under different discharge voltages was randomly selected to train GRU. The termination condition was set as 100000 trainings or the maximum error less than 0.15 mJ. The maximum error of model under each high voltage was less than 0.15 mJ (Fig. 6). Since the energy center value was 10 mJ, the relative error was less than 1.5%. The change of the maximum error in the training process indicates that the GRU neural network converges during the training process (Fig. 7). The data outside the training set was used to validate the model. The model obtained by training was used to calculate the laser light energy when the high voltage was 1550 V, and the comparison between the obtained energy value and the energy value collected on the actual laser after processing (1) is shown in Fig. 8. The energy obtained through the GRU neural network has a good coincidence with the energy of the actual pulse. Another verification data set was collected at laser working with repetition frequency of 1, 2, 3, and 4 kHz. The maximum error between the model data and the actual laser data was less than 0.13 mJ under different repetition frequencies, that is, the relative error was less than 1.5% (Fig. 10).

**Conclusions** The energy model of excimer laser is a complex nonlinear model, which is difficult to get an accurate model from the theory. However, the actual research and application work need an accurate laser output energy model. In this paper, through the method of deep learning, GRU neural network was to identify the energy model. The verification results show that the maximum error between the pulse energy generated by the laser energy model identified by GRU neural network and the actual energy was less than 1.5%. The maximum error 1.5% is less than 2.74% of the required energy stability in dose control, which meets the simulation requirements of the model control effect. This method can accurately identify the laser energy model. Using the identified model can be more convenient for the simulation of energy control algorithm, so as to improve the energy stability control and dose accuracy control of excimer laser.

**Key words** lasers; gate recurrent unit network; excimer laser; model identification

**OCIS codes** 140.2180; 140.3425; 140.3538