Journal of Crystal Growth 583 (2022) 126527

Contents lists available at ScienceDirect

Journal of Crystal Growth

journal homepage: www.elsevier.com/locate/jcrysgro

Evaluation of crystalline structure quality of Czochralski-silicon using near-infrared tomography

Mathias N. Jensen, Olav Gaute Hellesø *

Department of Physics and Technology, UiT The Arctic University of Norway, Hansine Hansens veg 18, Tromsø 9019, Norway

ARTICLE INFO

Communicated by Thierry Duffar

Keywords:
A2. Czochralski method
A1. Tomography
A1. Crystal structure
A1. Defects

ABSTRACT

In this work, three silicon samples are subject to tomographic scans using a 1.6μm laser. The samples were prematurely terminated due to anomalies during the Czhochralski-process. They are taken as analogues of the in situ crystal, where one sample has known aberrant structure in its lowermost 45 mm. The results of the tomographic scans show a distinct difference in transmission profile between the material of known poor mono-crystalline structure and assumed good structure. Three different analysis tools are constructed and applied to quantify the quality of the structure from the results of the tomographic scans. The first two analysis tools are applied as correlation filters constructed from patterns resembling the indicative transmission profiles of high-quality structure, one pattern being an ideal square wave and the other being experimentally determined from the measurements. Both correlation filters yield clear differentiation of low- vs. high-quality material. The final analysis tool is a deep convolutional neural network (deep CNN) evolved from a predetermined architecture configuration using a genetic algorithm. The trained CNN is shown to differentiate the usable high-quality material from the unusable material with a 98.7% accuracy on a testing set of 76 profiles and successfully assigns quality factors to the material that are in good agreement with the correlation filters and previous observations.

1. Introduction

In the production of silicon crystals, a mono-crystalline ingot, or boule, is "pulled" from a crucible of molten material. This process relies on strict control of the thermal conditions surrounding the growth interface as even small deviations can cause the formation of anomalies, such as crystal dislocations or, in extreme cases, a complete loss of mono-crystalline structure[1]. This results in the material forming intersecting crystal lattices of various sizes and orientation, producing material with non-homogeneous macroscopic properties. As the goal of this type of growth process is to obtain high quality material with a homogeneous structure throughout, such deviations are detrimental for the outcome. For an intact mono-crystalline structure, four ridges parallel to the growth axis can be observed along the crystal sidewall. These ridges, or nodes, are caused by the cubic nature of the silicon crystal and can be used as indicators for an intact structure. Inversely, their disappearance can also be taken as the indicator for a loss of mono-crystalline structure[2]. However, this can only be observed after the structure has been lost throughout the cross section of the crystal, making it poorly suited both for determining the precise point where the usable mono-crystalline material ends and for predicting the occurrence of structure loss. This work aims to explore the use of tomographic scanning through the center of such crystals using a near-infrared laser to detect variations in transmission related to crystal abnormalities such as dislocations or aberrant structure. Three crystal samples that have been prematurely separated from the melt due to process anomalies are used. One of the samples, having a clear loss of structure, has been shown in a previous work[3] to have an anomalous transmission vs. angle profile when compared to the other two. The transmission vs. azimuth profiles through the aforementioned samples are obtained over a range of scan heights and analysed using three pattern recognition methods to provide a measure of the quality of the mono-crystalline structure (degree of homogeneity) and determine the starting point of the unusable material (lost macroscopic structure). The outcome is assessed for quality assurance applications and as an in situ monitoring tool during the production of mono-Si in the Czochralski-process.

* Corresponding author.
E-mail addresses: mathias.n.jensen@uit.no (M.N. Jensen), olav.gaute.helleso@uit.no (O.G. Hellesø).

https://doi.org/10.1016/j.jcrysgro.2022.126527
Received 10 November 2021; Received in revised form 5 January 2022; Accepted 9 January 2022
Available online 22 January 2022

2. Hypothesis and expectations

Given that pure silicon has an intrinsic band gap energy of $1.12eV$ at $300K$[4], no direct band-to-band electron transitions can occur for photon energies lower than $1.12eV$, and therefore the material is expected to be highly transparent for wavelengths exceeding $1.1\mu m$[5]. Because of this, a laser of wavelength $1.6\mu m$ ($0.775eV$) is used in this work. However, as the material of the samples is extrinsic in nature, absorption can still occur through electron transitions to and from free carrier states induced by the presence of the dopant (Boron). The magnitude of this absorption is given from the work of Schroder et al. [6]:

$$\alpha = \frac{q^3\lambda^2 p}{4\pi^2\varepsilon_0 c^3 nm^{*2}\mu} \approx 2.7{\cdot}10^{-18}\lambda^2 p \tag{1}$$

The dopant concentration specified by the manufacturer is $1.45{\cdot}10^{16} \geqslant p \geqslant 2.70{\cdot}10^{16}$ atoms pr. cm^3 for a main section positions $0 \geqslant z \geqslant 1500mm$. Assuming a constant diameter of $200mm$, this gives an expected absorption of $-13 \pm 4dB$ for a full length sample. Accounting for the reflective loss, assuming normal angle of incidence and a refractive index of 3.44[7], the expected transmission is $-16 \pm 4dB$ for the samples used in this work.

As explored in a previous work[3], the geometry of the protruding ridges of the nodes is expected to obscure the beam of the laser, resulting in no detected transmission for the azimuth of the nodes. In the case of high-quality mono-crystalline structure, it is expected that the transmission should remain mostly constant between the nodes, producing a transmission cross-section similar to what is shown in Fig. 1. In the findings of the previous work, it was also noted that while samples with "good" overall structure (highly ordered) yielded profiles that corresponded well with the expectation, a sample with known "bad" overall structure (high defect density) yielded a very different profile. In the regions where the mono-crystalline structure was compromised, the profile exhibited sparse, but intense peaks, in lieu of flat transmission in the "clear" sections between the node azimuths. The profile was shown to have eight dominant peaks at azimuths leading and lagging the four nodes by approximately 15° for the low-quality material, in contrast to the four wide plateaus observed for high-quality material[3]. From the work of Meyer et al.[8], it is known that the slip planes between homogeneously structured lattices exhibit highly anisotropic scattering, resulting in a "smearing" effect on the transmitted radiation image (see Fig. 7 of Meyer *et al.*[8]) with a scattering direction parallel to the slip plane. It is postulated that the eight distinct peaks in the profile are artefacts produced by a similar effect except that there are, in our case, multiple slip planes formed by the interfaces of the pseudo-randomly oriented lattices that form after a loss of mono-crystalline structure. From this, it is hypothesized that this abnormal transmission profile may be used as an indicator for the presence of such chaotic structure and may therefore be used as a measure of the quality of the mono-crystalline structure in cylindrical silicon ingots.

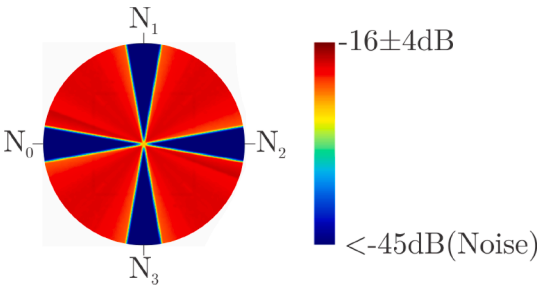

Fig. 1. Cross section of idealized transmission profile shown with nodes (N_{0-3}).

3. Experimental method

3.1. Samples

The samples shown in Fig. 2, conforming to the same specification and originating from the same foundry, are p-type (Boron doped) crystalline silicon boules manufactured to a specified diameter of 200 mm, with their dimensions, mass and bottom surface deflection (h) listed in Table 1. All three samples are the result of premature separation from the melt, resulting in the growth process being abruptly terminated such that they are effectively "snapshots" of the crystal during growth, and may therefore be used as analogues for the material in situ.

While all samples are expected to have some anomalies due to the separation shock, sample 1 shows clear indication (node termination) of unexpected loss of structure. These indications are not present in neither of two other samples. The loss of structure in sample 1 is evident by the absence of node lines in the lowermost 45 mm of the sample, see Fig. 3, and is supported by the presence of slip lines at varying positions surrounding the node lines higher up on the sample, see Fig. 3. Sample 1 is thus taken as an example of having "bad" material in its lower section, while samples 2 and 3 are taken as having mostly "good" material.

3.2. Experimental setup

The experimental setup is shown in Fig. 4, for a sample with lost structure similar to what is expected from sample 1. The central part of the setup consists of a rotating platform upon which the sample is placed. The platform is belt-driven and actuated by a DC motor providing a total gear ratio of 1:180. The laser (New Focus, Velocity set to $1600nm$) is mounted on a vertically oriented motorized translation stage (Standa, 8MT50-150BS) with a half-wave plate (Thorlabs, WPLH05M) on its output to control the beam polarization. A detector (Electro-Optical Systems, IGA-010-TE2-H) is mounted on an identical translation stage (Standa, 8MT50-150BS) such that it can travel parallel to the laser over a range of $140mm$ along the z-axis. Due to the small active area of the detector ($1mm$ diameter), a focusing lens (L) is added, increasing the pupil diameter of the detector to $25.4mm$ with an NA of 0.03. Due to the high and fixed sensitivity of the detector amplifier ($9{\cdot}10^7 V/W$), an ND-filter (effective OD 3.0 for $\lambda = 1.6\mu m$) is added to avoid saturation. This gives a high signal-to-noise ratio and matches the detectable intensity range with the observed transmission intensity, allowing better use of the detector range.

The signal from the detector is collected using a DAQ (National Instruments, USB-6009). An ad hoc timing system is implemented to serve as feedback for the DC motor, feeding a clock signal to the DAQ with a frequency of one pulse per rotation. The speed of the DC motor is manually set using a constant voltage source, while control of the

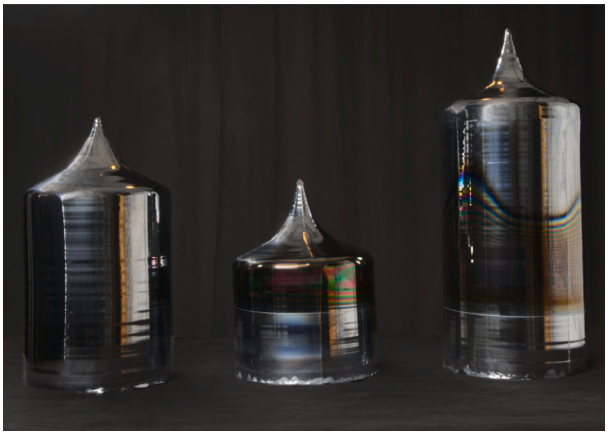

Fig. 2. Picture of silicon samples 1–3, from the left.

Table 1
Sample specifications.

	Length (*mm*)	Diameter (*mm*)	Mass (*kg*)	*h* (*mm*)
Sample 1	276 ± 1	213 ± 1	23.7	19.7 ± 0.1
Sample 2	162 ± 1	211 ± 1	10.1	11.0 ± 0.1
Sample 3	389 ± 1	212 ± 1	34.4	13.8 ± 0.1

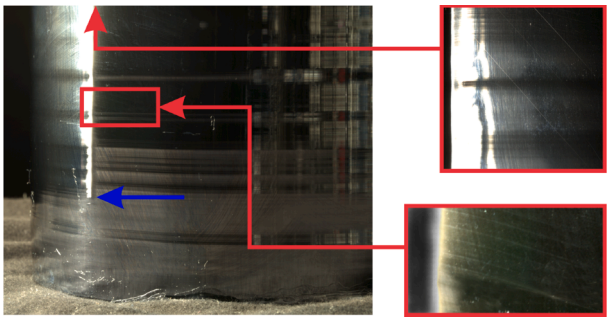

Fig. 3. Picture of sample 1 showing node termination (blue arrow) and slip lines (red frames).

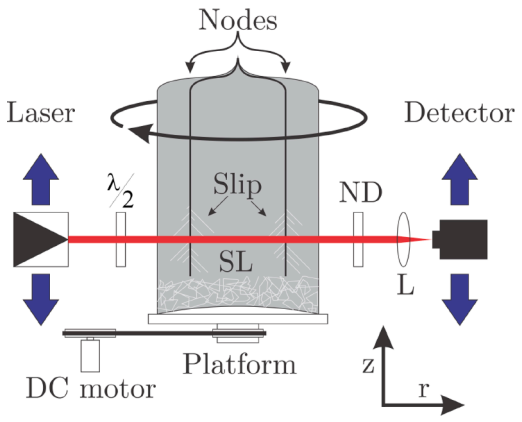

Fig. 4. Outline of setup for measuring a sample with macroscopic dislocations (slip lines) and complete collapse of mono-crystalline structure (SL).

translation stages, as well as data capture and processing, is automated (Python, V3.9.4 X64).

3.3. Scanning procedure

The scanning procedure is initiated by a mapping of optimal alignment for a predetermined set of scan positions (z-axis). This is conducted by setting the laser position to a predetermined scan position and moving the detector z-position in a range $\pm 5mm$ about the laser position.

For each detector position, data is acquired over a set number of rotations (1 by default), which is analyzed and assigned a score value. This score is determined as the mean of the upper 25th percentile of the data, encouraging the finding of a high transmission alignment, divided by the standard deviation in the percentile, discouraging the finding of an alignment that favours one or few peaks. The optimal alignment associated with each laser position is then set as the detector position with the highest score value.

Once a map of the optimal alignment positions has been obtained for a sample, the tomographic scan begins. The algorithm acquires a data set from each of the scan positions, henceforth referred to as a "slice". Each slice consists of data from a set number of revolutions (3 by default) acquired at maximum sampling frequency (24 kHz). The ad hoc timing system is used to ensure that each data-point in the slice is accurately mapped to the true azimuth at the time of acquisition. This is performed by using the clock pulse from the ad hoc timing system to confirm the rotation frequency of the crystal and setting the acquisition time accordingly. Using the clock pulse as a fixed reference for the acquisitions ensures that all slices begin at the same azimuth, and thus use a common coordinate system. To maintain accurate tracking of the rotation frequency and azimuth, the timing system periodically re-calibrates after a set number of revolutions (10 by default), in addition to the initial calibration.

4. Experimental results

The lowermost section of the samples include the concave bottom surfaces, with deflections listed in Table 1. No transmission is expected due to reflection by the bottom surface[3]. Therefore, a starting scan height of $10.5mm$ is selected. All three samples are scanned at 166 z-positions with a resolution of $0.5mm$ from $10.5mm$ to $35.5mm$ above the separation plane (bottom edge), and a resolution of $1mm$ from $35.5mm$ up to $150.5mm$.

As depicted in Fig. 5, the transmission profiles for samples 2 and 3 (Figs. 5b and c, respectively) remain largely constant throughout the samples, as expected. However, sample 1 exhibits a very different profile, that changes drastically from the sparse, sharp peaks for the lower slices to a wider plateau for the highest slice, approaching the profiles seen in samples 2 and 3. In terms of total transmission, the peaks of the profiles through samples 1 and 2 appear to remain in the vicinity of $-20dB$, which is at the lower end of the expected range. Transmission through sample 3 is somewhat lower than for the other two samples, with peaks around $-23dB$. While the absolute intensity is lower than expected, the signal-to-noise ratio is high, at $25 \pm 2dB$, and the profiles of the slices clearly distinguishes high-transmission azimuths from low-transmission azimuths.

The images shown in Fig. 6 are reconstructions of the transmission map through the crystals, obtained from the data presented in Fig. 5, using a simplified inverse Radon transform with one data-point per azimuth. The intensity of the images is given by the magnitude of the transmission for a given azimuth, but is subject to normalization and Gaussian smoothing on a per-slice basis. The images thus show the relative transmission throughout the cross section, separate from the absolute scale shown in Fig. 5.

As can be seen in the top row of Fig. 6, all three samples produce patters similar to the predicted pattern illustrated in Fig. 1, showing four clear sectors with relatively high and uniform transmission. In the next row down, this is no longer the case, as sample 1 exhibits sparse peaks with azimuths approximately symmetric around the nodes. Both samples 2 and, especially 3, continue to display a transmission profile similar to the predicted profile. In the lowermost row, the same dichotomy is present, albeit the signal-to-noise ratio for all samples is reduced. As previously stated in Section 3.1, sample 1 shows clear evidence of abnormal crystalline structure, both from the termination of the nodes and the presence of slip lines as shown in Fig. 3. From this knowledge and the observed patterns in transmission for the slices intersecting these areas, a correlation between quality of the mono-crystalline structure and the transmission profile is implied.

5. Data processing and analysis

5.1. Correlation filtering

The simplest method of differentiating two underlying patterns is through the use of correlation filtering by using a target pattern (i.e. transmission profile of high-quality material) and determining the conformity of a separate pattern to the target (i.e. quality factor) by the cross-correlation of the two. As we are interested in a scalar value for this, the peak cross-correlation between the target and measured profiles

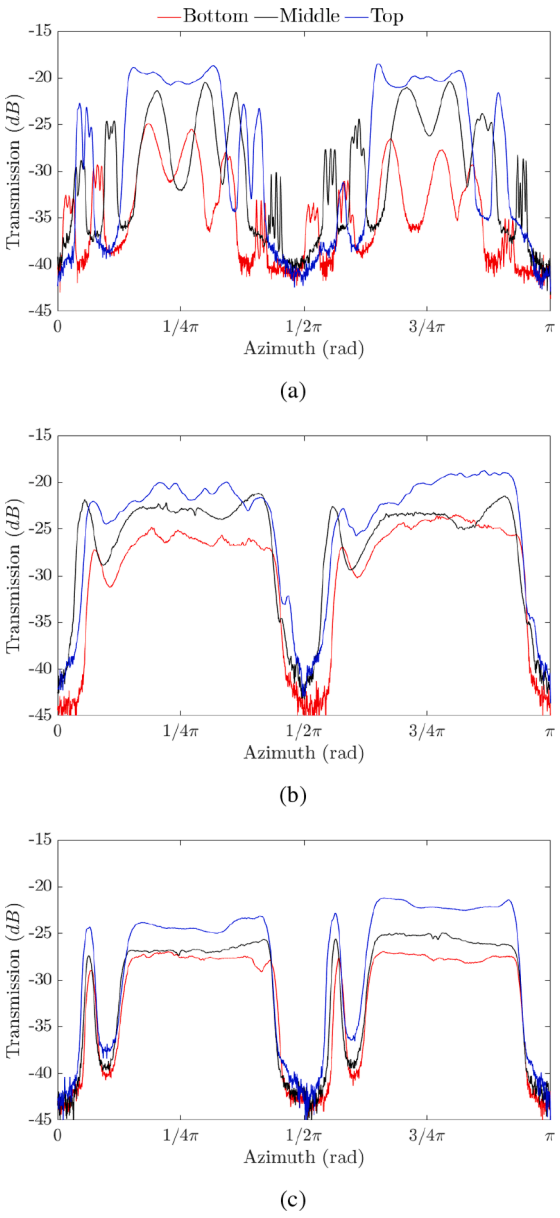

Fig. 5. Absolute transmission through the three samples at the lowermost (Bottom), intermediate (Middle,$z = 35.5mm$), and the uppermost common height (Top, $z = 140.5mm$). (a) Sample 1 (b) Sample 2 (c) Sample 3.

are taken as the raw output. The raw output of all slices considered are then normalized to a range $[0, 1]$ to make them easily comparable.

Correlation filtering is computationally inexpensive, especially in the case of a single dimension as here, which allows for easy integration as a real time feedback mechanism. A significant disadvantage, however, is the low order of complexity and thus limited ability to model complex patterns. Another shortcoming of such filters is that they are sensitive to background levels, scale and noise/distortions. These factors must therefore be attenuated by pre-processing the input passed to the filter, but the effectiveness of this is limited and often relies on assumptions regarding the properties of the factors.

5.1.1. Ideal filter

As previously stated, the transmission profile through a high quality crystal is expected to form two wide, pseudo-flat plateaus that are well distinguished from the underlying noise and background, as is the case throughout samples 2 and 3 (see Figs. 5 and 6).

The obvious candidate for modelling such a profile is a square wave

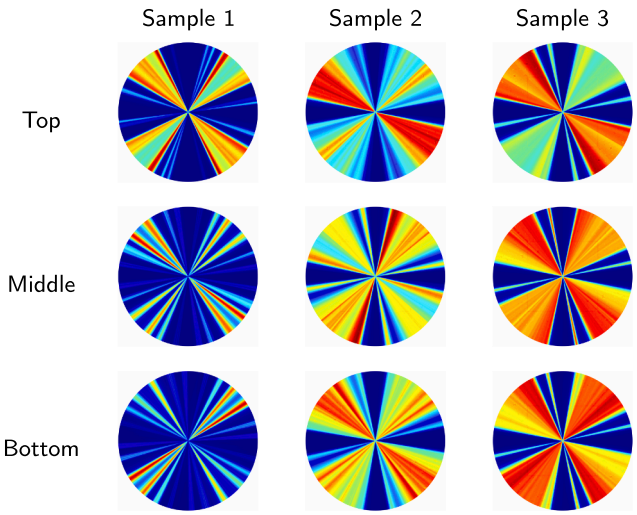

Fig. 6. Reconstruction of transmission cross-sections obtained from the measurements, with normalized linear scale indicating dark red as maximum transmission.

as shown in Fig. 7a, producing a slice cross-section as shown in Fig. 1. From this target profile, a filter is constructed and applied to the data as a correlation filter, producing the results shown in Fig. 7b. It is clear that the filter responds quite well to samples 2 and 3, returning a relatively

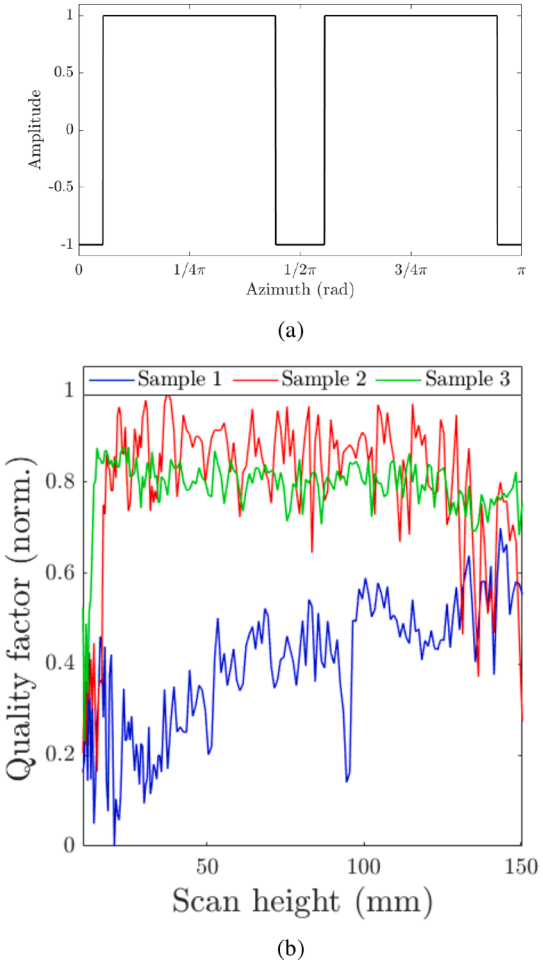

Fig. 7. Ideal correlation filter used to determine crystal quality. (a) Target profile (b) Filtering result.

high and constant value for all slices above the interface peak. Contrary to this, the filter responds relatively poorly to sample 1, returning a low response that increases towards that of samples 2 and 3 as the scan position increases. The response of this filter agrees well with the hypothesis and the observations made previously regarding the condition of the crystalline structure in sample 1 versus that of samples 2 and 3. The results of this filtering are encouraging and differentiates the known high-quality structure from the known low-quality structure with good separation of the filter response of sample 1 from that of samples 2 and 3. However, the filter response, and thus assigned quality factor, contains a high degree of noise, especially for samples 1 and 2, introducing uncertainty in determining whether the material is of usable quality or not.

5.1.2. Experimentally obtained filter

Another target profile is constructed based on the experimentally determined profiles of the known high-quality material of samples 2 and 3. The target profile is defined as the mean profile of all slices above the apex of the bottom surface (transmitted beam is detected) for samples 2 and 3. The target profile is shown in Fig. 8a normalized to $[-1,1]$. This filter is applied in the same manner as the ideal filter and the result of the correlation filtering is shown in Fig. 8b. The filter response of the experimental filter to the data is highly similar to the response of the ideal filter in regards to its differentiation of sample 1 from samples 2 and 3. However, it appears to give a slight reduction in the noise of the response, but also gives slightly poorer separation of sample 1 from samples 2 and 3. In contrast to the ideal filter, the experimental filter better illustrates the change in transmission profiles in the slices of sample 1 as the scan height increases, showing a gradual increase until it merges with samples 2 and 3. However, in terms of differentiating the high-quality and low-quality structure, the experimental filter shows inferior performance compared to the ideal filter as the separation between sample 1 and samples 2 and 3 is small compared to the noise, giving an unfavourable overlap between them.

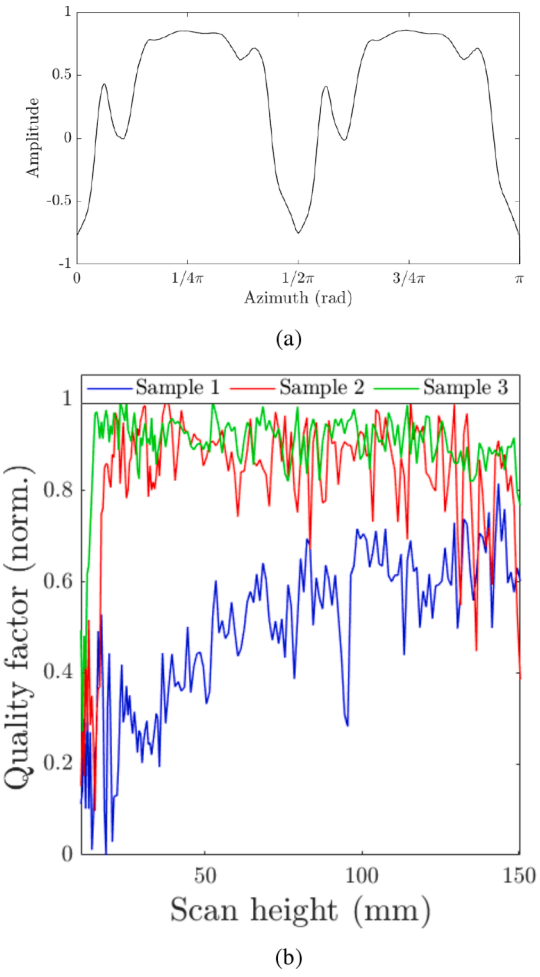

Fig. 8. Experimental correlation filter. (a) Target profile (b) Filtering result.

5.2. Neural network

As the underlying mechanisms causing the difference in the profiles of sample 1 from samples 2 and 3 are currently not well known, the features of the profiles may include complexities that are not immediately evident. While the results from both the ideal and experimental filter illustrate a clear contrast between the lower sections of sample 1 and the remaining slices, these filters may be overly simplistic and the results may therefore be, to a degree, circumstantial.

An alternative to the use of a predetermined target profile for correlation filtering is to design a filtering mechanism that adapts itself to the observed data, namely a machine learning model. The most logical type of machine learning model to use for this kind of analysis is a convolutional neural network (CNN).

The use of models such as CNNs, and other variants of neural networks[9], is becoming increasingly popular in the field of data analysis, especially for patterns that are not well known or subject to distortion/noise. The driving reason for this is that, contrary to the stiffness of correlation filters, CNNs are highly flexible and can learn to adapt to a wide range of patterns and features. Deep convolutional neural networks, with their increased depth (number of layers), are also able to learn highly abstract and complex patterns, making them able to model any pattern (given sufficient depth and width) and can learn to adapt to high degrees of distortions/noise in the data.

However, contrary to correlation filters, neural networks, and especially deep networks, rely on a longer sequence of operations that involve many more parameters, often several million, rather than the single operation of correlation filtering. Neural networks must also be taught how to determine its output, which requires a sizable pool of examples with known attributes (quality factor) and a time-consuming training phase, before it can be implemented. This makes then both computationally expensive and requires data with known properties to learn from, making them challenging to implement in real-time.

5.2.1. Architecture

A CNN consists of two main sections, the convolutional sub-network and a fully connected sub-network. The convolutional sub-network usually contains many, relatively small, learned filters and, as the name suggests, forms an output by discretely convolving the input with the filter. Such networks usually have many layers of such filters in series to form an abstract representation of the input and each layer commonly uses multiple independent filters in parallel to produce multiple output channels containing different representations of the input. A CNN normally uses this convolutional sub-network as a preprocessing mechanism that reduces the input to a concentrate which is then fed to the fully connected (FC) sub-network. The fully connected sub-network is a classical neural network consisting of a determined number of layers (depth), each with a determined number of neurons (width). After the input has passed through the fully connected sub-network, it reaches the head of the network, which produces the final output of the network based on the output of the last layer of the fully connected sub-network. The architecture of the chosen deep CNN is shown in Fig. 9.

As the input to the network is a one dimensional vector, all of the convolutional filters are also vectors. As shown in Fig. 9, this architecture also employs connections that directly bypass the filters, so-called ”skip-connections” common to the ResNet architecture[10]. The use of skip-connections encourages the network to learn filters that modify the input to produce the output, rather than creating an entirely new representation. This added bypass helps the network learn faster by

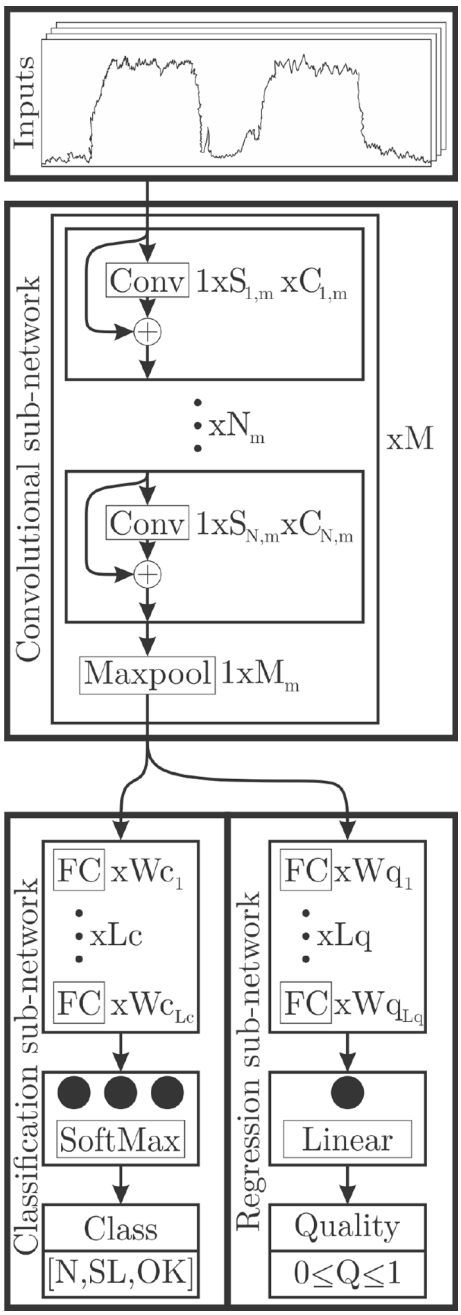

Fig. 9. Generalized architecture of deep CNN with modular convolutional blocks, taking the transmission profiles as its input and returning a quality factor and prediction of the usability of the material (OK vs. N or SL).

smoothing out the loss surface[11], allowing the gradient descent algorithm to more easily converge to the optimal network parameters.

The overall structure of the network is designed to be highly modular, by grouping the layers into blocks that can be configured as needed. The convolutional sub-network contains M convolutional blocks, each terminated by a maxpool layer to reduce the size of the block's output. For each block $m \in [1..M]$, there are N_m convolutional layers, and each layer $n \in [1...N_m]$ has a defined filter size $S_{n,m}$ and number of parallel filters $C_{n,m}$ that can be configured individually, as needed.

Once the input has been passed through all convolutional blocks, it is returned as the output of the convolutional sub-network and passed to two parallel, fully-connected sub-network. The two sub-network are independent and may be configured separately with any chosen depth (Lc and Lq) and width (Wc_k for $k \in [1..L_c]$ and Wq_l for $l \in [1..L_q]$). The classification sub-network aims to learn a way of predicting if the input originates from a noise-only signal (N), the profile of a slice with low-quality structure (SL), or the profile of a slice with high-quality structure (OK). This is achieved using three terminating neurons that take the final layer of the classification sub-network as input, weighting them and summing them before passing the result to a softmax activation function that redefines the raw output of the neurons into a probability distribution. The predicted class is then taken as the class with the highest probability given by the softmax output. The regression sub-network assigns a numerical value to the quality of the structure, based on the input profile, to create an output that is comparable with the filter responses shown in Figs. 7b and 8b. The output of this sub-network is determined as the weighted sum of the last layer in the sub-network and returned as a number (Q).

5.2.2. Configuration and evolution

Given that the presented data is in the form of single-dimension vectors, a small network should be sufficiently detailed, but the balance between the size and complexity of the convolutional, classification, and regression sub-networks must be chosen carefully. Due to the unknown complexity of the underlying features and their distribution in the data set, a larger network with many parallel convolution filters may provide greater performance than a small one. However, such a large network would require more example profiles to learn from and may be prone to over-fitting to the data without learning the underlying pattern. Conversely, a small network with more convolutional layers, but few parallel filters, would be superior in forming abstract patterns due to the extensive transformation of the input through many convolutional layers. This may help the network learn the underlying pattern better, especially on small training sets, but the lack of parllel representations of the input may limit the level of complexity the network can account for.

Because of this challenge, further inspiration is taken from nature by introducing the concept of evolution through a genetic algorithm[12], loosely based on the method employed by Dahou et al.[13]. This algorithm defines 13 hyper-parameters ("genes") that define the structure of the networks, these then define:

- Number of convolution blocks and layer configurations within each block
- Size of convolution filters in each block
- Number of parallel filters in each block
- Width and depth of classification and regression FC layers, separately
- Drop out rate for convolution, regression, and classification sub-nets separately
- Activation function for convolution, regression, and classification sub-nets separately

This algorithm then creates a population of five networks from a predetermined pool of configuration hyper-parameters using a random combination of the hyper-parameters. These five networks (generation zero) are then partially trained over 10 epochs on a common training set before evaluation on a separate common test set, defining the fitness score as the inverse of the loss (cross-entropy). To prevent devolution, the fittest network is always passed to the next generation such that the best of every subsequent generation is always as good or better than the previous one, driving the "survival of the fittest"-mechanism. The top two networks of the previous generation are then hybridized to produce two new networks. This process randomly combines the hyper-parameters ("genes") of the "parents" to produce "offspring" with a configuration containing only those "genes" that prove most advantageous, while providing a mechanism for mutation. The remaining two slots of the next generation are filled by producing two random networks from the original pool of configurations and hybridizing them with the two top performing networks from the previous generation. This is done to increase the genetic diversity of the population, increasing the

likelihood that advantageous "genes" outside those of generation zero are introduced into subsequent generations. After six generations of evolution, the process is terminated and the best performing network is taken as the "apex"-network, giving the configuration listen in Table 2.

In addition to the evolved configuration shown in Table 2, the convolution sub-network employs a rectified linear (ReLU) activation function with dropout rate 0.4, the classification sub-net employs a ReLU activation with dropout rate 0.5, and the classification sub-net employs a sigmoid activation with dropout 0.3.

5.2.3. Training and results

The original data set used to generate Figs. 7b and 8b contains 166 profiles for each of the three samples (498 in total). This is not sufficient to train the CNN, especially since the true nature and quality of the material in sample 1 is unknown. Therefore the data set is augmented by adding profiles obtained from earlier phases of the experiment, when polarization sensitivity was investigated. These profiles showed no discernible sensitivity to polarization but are taken at different scan heights from the final phase. Thus, they are not evaluated in parallel to the main set, but can be considered to represent the same patterns, bringing the total number of profiles to 996. Due to the true nature and quality associated with certain profiles from sample 1 being unknown, these are omitted, reducing the number of profiles viable for training to 675. To encourage the network to learn the underlying pattern instead of specific features of the profiles, the data set is augmented by creating synthetic profiles from the measured real profiles. These synthetic profiles are randomly selected from the expanded set of 675 profiles and a random selection of these are again modified with added Gaussian noise (mean of zero and standard deviation $-13dB$ of profile amplitude) to emulate measurements with a lower signal-to-noise ratio. Since the transmission for a given azimuth is assumed to be independent of preceding measurements, a change in direction of rotation is emulated by time-reversing (reversing the mapping of data-points to azimuth) a set of randomly selected profiles from the expanded set. The augmentation generated 508 synthetic profiles (75.3% of real profiles), giving 1183 profiles usable for training the CNN.

The profiles in the data set are then randomly shuffled to homogenize the set, increasing the likelihood that the subsets of it used by the CNN during training and testing are representative of the whole. The profiles in the set are also given a class label and a quality label for the CNN to use as reference during training. As the exact numerical value of the material quality is not well known, noise is added to the quality labels to encourage the model to learn a more abstract quantification rather than a discrete binary one.

The shuffled and labeled data set is then randomly separated into a testing and a validation set containing 10% of profiles each, while the remaining 80% is reserved for training.

As training during the evolution phase uses the same source data as here, there is a risk of overlap between the training set used during evolution and the testing or validation sets used here, which would bias the results. To prevent this, the "apex"-network is rebuilt such that its hyper-parameters ("genes") are kept, but all parameters learned from training are reset.

The "apex"-network is then trained on the generated training set over a total of 40 epochs using an Adam optimizer with a learning rate reducing from 10^{-2} to 10^{-5} over four steps (10 epochs per learning rate). The regression sub-network is subsequently trained on top of the main network, using the same training data and procedure.

Once trained, the performance of the network is evaluated using the testing set, giving an overall prediction accuracy of 92.2% and the confusion matrix shown in Table 3. The CNN predicted 98.7% of the known "good" material profiles (*OK*) as "good" ($\widehat{OK}$) while none of the known "bad" (*SL*) or pure noise (*N*) are misclassified as "good". However, there is some misclassification between the pure noise and the "bad" material as shown by 38.5% of the true "bad" being misclassified as noise and 13.3% of true noise being misclassified as "bad" material. The pure noise case is only present for prematurely separated samples (due to the concave bottom surfaces) and not in a complete boule. Thus, the important factor is the overlap between the "bad" and "good" predictions, which can be argued to be taken as less than or equal to the 1.3% overlap between *OK* and $\widehat{N}$.

While the performance of the network in differentiating the noise only (*N*) from the "bad" structure (*SL*) is not optimal, its capability in differentiating the unusable (*N* or *SL*) from the usable material (*OK*) is shown to be excellent with one misclassification among 75 true good profiles. Upon investigation, it is found that the misclassified profile is a weak signal with artificially added noise, making it appear as pure noise. It can also be observed that the CNN responds to the synthetic and real data in the same manner, predicting the same class and approximately equal quality for both real data and the derived synthetic data. This confirms that the synthetic data contains the same underlying pattern as is present in the real data and that the CNN is able to recognize the pattern in the presence of modifying factors such as noise or a change in direction of rotation.

Applying the network to the data set presented in Figs. 7b and 8b yields the classification shown in Fig. 10a and the quality factor shown in Fig. 10b. As seen in Fig. 10a, the network returns a high degree of certainty that there is no bad structure in samples 2 and 3, but that there is a high density of such occurrences in the lower section of sample 1. The quality factor shown in Fig. 10b, agrees with the results of correlation filtering shown in Figs. 7b and 8b, but shows better separation of the low-quality material of sample 1 and the high-quality of samples 2 and 3, while also exhibiting significantly less noise. Both the quality factor and the classification of "good" vs. "bad" structure agrees well with what is inferred in Section 2 regarding the quality and viability, of the material in sample 1 increasing with distance from the onset of structure loss.

5.2.4. Robustness of the trained model

The purpose of using a CNN instead of the simpler forms of filtering, such as described in Section 2, is its lower sensitivity to noise in the input signal. To illustrate this, the trained CNN is fed two previously unseen profiles, one with known low quality and one with known high quality, while observing the class (N, SL or OK) and quality factor predicted by the CNN.

The two profiles are replicated 100 times and contaminated with randomly generated noise before being passed to the CNN, the robustness of the network is then evaluated by increasing the severity of the added noise until the output of the CNN is affected by either a change in predicted class (ΔC) or a 5% change in assigned quality-factor(ΔQ). For fairness, three types of noise are used: Gaussian ($\mathcal{N}(0,\sigma)$), Poisson ($A{\cdot}\mathcal{P}(\lambda)$), and noisy sine ($B{\cdot}Sin(\propto 1/p) + \mathcal{N}(0, B/10)$). These noise types

Table 2
Configuration of CNN.

Block	Layers	Size	Channels	Parameters
Conv 1	5	5	32	21 k
Conv 2	5	5	64	92 k
Conv 3	5	5	128	369 k
Conv 4	5	5	256	1.5 M
FC Class	5	700	-	22.2 M
FC Regr.	6	512	-	16.1 M
			Total	40.3 M

Table 3
CNN confusion matrix for test data set.

	$\widehat{N}$	$\widehat{SL}$	$\widehat{OK}$
N	86.7%	13.3%	0.0%
SL	38.5%	61.5%	0.0%
OK	1.3%	0.0%	98.7%

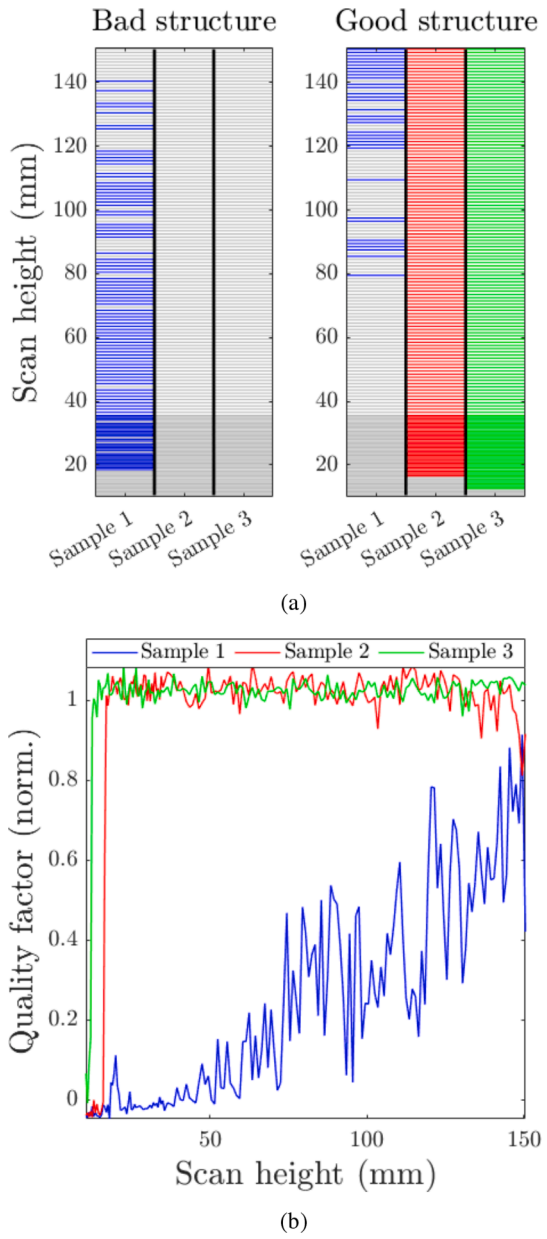

Fig. 10. Result of deep learning filter. (a) classification of low- vs. high-quality predictions (b) Assigned quality factor of material vs. slice height.

are then tried for all combinations of their parameters (1000 points pr. parameter) and are considered for the extremes of the noise types, returning the signal-to-noise thresholds listed in Table 4. As shown from the noise testing, the CNN is quite robust against all three types of noise, showing robust predictions for input profiles with an induced signal-to-noise ratio down to $5dB$, compared to the measured SNR of $> 23dB$ as mentioned in Section 4.

The regression sub-network is shown to be more robust than the classification sub-network, being able to withstand $1.4-9.9dB$ lower SNR without deviating more than 5% in predictions. Notably, it is also observed that all the noise-induced misclassifications are on the low-quality profile while the high-quality profile is not misclassified, implying a higher false positive rate than false negative rate. Albeit not shown, both classification and regression outputs of the CNN exhibit a slightly higher ($1.0-1.7dB$) tolerance to sinusoidal noise if the period of the noise is a whole fraction of a rotation ($360°, 180°, 120°…36°$).

Table 4
CNN noise threshold.

Noise type	Parameter	Threshold	ΔC	ΔQ
Gaussian	-	σ^{-1}	$2.6dB$	$-5.2dB$
Poisson	$\lambda = 0.2$	A^{-1}	$-7.0dB$	$-16.9dB$
	$\lambda = 20$	A^{-1}	$-6.3dB$	$-7.7dB$
Sine	$p \in [0, 10]$	B^{-1}	$5.0dB$	$-7.7dB$

6. Conclusion

Three samples of mono-crystalline silicon with diameters of $212 \pm 2mm$ are observed to have a transmission of $-21 \pm 2dB$ for a near-infrared laser of wavelength $1.6\mu m$. Tomographic scanning is performed on all three samples over a lateral range of $140mm$ from $10.5mm$ to $150.5mm$ above the separation plane. A total of 498 transmission profiles, measured as a function of transmission vs. crystal azimuth, are obtained for the three samples.

The transmission profiles are shown to produce a consistent pattern for a given vertical slice (z-position) of the crystal. All recorded transmission profiles from all three samples exhibit the same four blackout-zones due to beam refraction/obstruction by the node geometries. The profiles are also shown to exhibit unique features for slices intersecting material of known defective crystalline structure that differs notably from the profiles for slices intersecting intact mono-crystalline structure.

A quality-score can be determined from the transmission profile of a slice using targeted correlation filtering. An idealized square wave, with negative amplitude around the node azimuths, is used as the target profile to assign a quality factor to the measured profiles to produce a quality factor that is in agreement with observations and expectations regarding the samples. A second target filter is obtained from the measurement as the mean profile of assumed intact material from two of the samples. This filter also produced a viable quality factor in the presence of suspected poor/no structure, albeit with a higher degree of overlap between the known poor- and high-quality material compared to the idealized square wave filter.

A deep convolutional neural network (CNN) is also investigated as an analysis tool. The CNN is implemented as a modular architecture with a configuration determined by a set of 13 hyper-parameters that define the properties of each of its elements. The hyper-parameters of the CNN are determined by a process of evolution using a genetic algorithm to create a CNN whose configuration is best suited to learning the patterns observed in the profiles. The evolved CNN is configured with 20 convolutional layers, preceding a fully connected classification head of five layers and a fully connected regression head of six layers, giving a total depth of $20+5/6$ layers and 40.3 million parameters.

The CNN is then trained over a total of 40 epochs and tested on a separate testing set of 118 profiles, achieving an accuracy in differentiating the assumed intact structure from the known defective structure (or noise) of 98.7%. The predictions of the CNN results in a quality factor consistent with the results of both correlation filters, albeit with significantly reduced noise and vastly improved contrast between low- and high-quality material. The CNN also successfully classifies all slices of all three samples, yielding a map of lost vs. intact structure that agrees well with observations and expectations regarding the samples.

The observations and results of this work show that a consistent pattern in the transmission profiles coincide with the state of the crystalline structure, and that this can be used to quantify the quality of the structure. As these experiments are conducted at room temperature, it is feasible that these methods could also work for complete boules under the same conditions. The non-destructive nature of this method allows for quality testing to be conducted on every produced boule without the need for slicing, and subsequent loss of material. The method also make it possible to determine the precise boundary between usable and unusable material, enabling smaller margins to be used when removing the unusable material, thus improving material yield.

Because the temperature of the material during production is significantly higher than for the conducted experiments, it is unknown if these methods can be adapted as a real-time feedback system. Further investigations on near/mid-infrared transmission through silicon at high temperatures ($\geqslant 1400K$) are required.

Validation of the method proposed in this work would require additional material samples and added measurements of these, presumably through destructive methods such as carrier density imaging and/or lateral photovoltaic scanning of the sliced samples, to provide a more detailed reference point for the assessment of the method and analysis. Future work on this concept would also include investigations of other deep-learning architectures, both more comprehensive evolved networks and known established architectures such as ResNet, VGGXX, and GoogLeNet.

Author contributions

The second author suggested the problem and the authors conceived the measurement methods together. The first author conducted the experimental work and the data analysis. The authors contributed equally on the written work.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors gratefully acknowledge financial support from the Research Council of Norway (No. 302333), and would like to thank Norwegian Crystals AS for supplying the samples and relevant reference information.

References

[1] A. Lanterne, G. Gaspar, Y. Hu, E. Øvrelid, M. Di Sabatino, Characterization of the loss of the dislocation-free growth during czochralski silicon pulling, J. Cryst. Growth 458 (2017) 120–128.
[2] Øyvind S. Sortland, E.J. Øvrelid, M. M'Hamdi, M. Di, Sabatino, Investigation of pinholes in czochralski silicon ingots in relation to structure loss, J. Cryst. Growth 510 (2019) 1–6.
[3] M.N. Jensen, O.G. Hellesø, Measuring the end-face of silicon boules using mid-infrared laser scanning, CrystEngComm 23 (2021) 4648–4657.
[4] V. Alex, S. Finkbeiner, J. Weber, Temperature dependence of the indirect energy gap in crystalline silicon, J. Appl. Phys. 79 (1996) 6943–6946.
[5] S.E. Aw, H.S. Tan, C.K. Ong, Optical absorption measurements of band-gap shrinkage in moderately and heavily doped silicon, J. Phys.: Condens. Matter 3 (1991) 8213–8223.
[6] D.K. Schroder, R.N. Thomas, J.C. Swartz, Free carrier absorption in silicon, IEEE J. Solid-State Circuits 13 (1978) 180–187.
[7] D. Chandler-Horowitz, P.M. Amirtharaj, High-accuracy, midinfrared (450cm-1–4000cm-1) refractive index values of silicon, J. Appl. Phys. 97 (2005) 123526.
[8] M. Meyer, M.H. Miles, T. Ninomiya, Some electrical and optical effects of dislocations in semiconductors, J. Appl. Phys. 38 (1967) 4481–4486.
[9] M. Avci, S. Yamacli, Neural network reinforced point defect concentration estimation model for czochralski-grown silicon crystals, Mathematical and Computer Modelling 51 (2010) 857–862. 2008 International Workshop on Scientific Computing in Electronics Engineering (WSCEE 2008).
[10] Z. Wu, C. Shen, A. van den Hengel, Wider or deeper: Revisiting the resnet model for visual recognition, Pattern Recogn. 90 (2019) 119–133.
[11] H. Li, Z. Xu, G. Taylor, C. Studer, T. Goldstein, Visualizing the loss landscape of neural nets, 2018. arXiv:1712.09913.
[12] J. Shapiro, Genetic Algorithms in Machine Learning, Springer Berlin Heidelberg, Berlin, Heidelberg, 2001, pp. 146–168. doi: 10.1007/3-540-44673-7_7.
[13] A. Dahou, M.A. Elaziz, J. Zhou, S. Xiong, Arabic sentiment classification using convolutional neural network and differential evolution algorithm, Computational Intelligence and Neuroscience 2019 (2019) 2537689.