Explainable machine learning for the analysis of transport phenomena in top-seeded solution growth of SiC single crystal

Yuto TAKEHARA*, Atsushi SEKIMOTO*, Yasunori OKANO*, Toru UJIHARA** and Sadik DOST***

* Department of Materials Engineering Science, Osaka University
1-3 Machikaneyamacho, Toyonaka, Osaka 560-8531, Japan
E-mail: asekimoto@cheng.es.osaka-u.ac.jp

** Department of Crystalline Materials Science, Nagoya University
Chikusa-ku, Nagoya 464-8603, Japan

*** Crystal Growth Laboratory, University of Victoria
Victoria, BC V8W 3P6, Canada

Received: 28 May 2020; Revised: 6 August 2020; Accepted: 30 September 2020

Abstract
Silicon carbide (SiC) is a power semiconductor used to supply and control the electric power source. Top-Seeded Solution Growth (TSSG) method is a promising technique for producing high-quality SiC single crystals. In order to achieve a high- and uniform-growth rate in this growth technique, however, the complex fluid flow developing in the growth melt/solution, mainly induced by the electromagnetic field of the induction-heating coils, free surface tension gradient, and buoyancy, must be well-controlled. Our previous studies have shown that the applications of a static magnetic field and seed rotation are effective in controlling the components of this melt flow and the associated control parameters were optimized effectively using the Bayesian optimization. In this study, we analyze the optimal state determined by the Bayesian optimization in more detail and it is found that the separation of the Marangoni flow near the seed edge leads to a non-uniform growth rate. In addition, the most sensitive region of the melt flow is determined by using an explainable machine learning technique based on a convolutional neural network and the sensitivity map obtained by SmoothGrad. This machine learning technique automatically predicts the preferred melt flow pattern that would lead to high-quality crystal growth. The interpretations by the explainable machine learning technique used in the present study are consistent with those of previous studies carried out on the optimization of the TSSG method.

Keywords : Bayesian optimization, Silicon carbide, Top-Seeded Solution Growth, Marangoni convection, Convolutional Neural network, Explainable machine learning, Sensitivity map

1. Introduction

Silicon carbide (SiC) is an attractive power semiconductor possessing favorable physical properties such as a wide band-gap and a high thermal conductivity. Top-Seeded Solution Growth (TSSG) method is a promising technique for producing high-quality SiC single crystals. In the TSSG method of SiC, solid Si is heated and melted in the graphite crucible by the induction-heating coils. The needed carbon dissolves from the crucible into the Si melt, then forms a dilute Si-C solution (melt). When the concentration of carbon is supersaturated near the single crystal silicon seed which is in contact with the melt, the growth of SiC single crystal begins on the seed.

There are several challenges in this growth technique. At first, a growth high temperature (about 2000°C) is required due to the low solubility of carbon into silicon. Higher solubility is necessary for a high growth rate. Secondly, since the growth system is opaque, the transport phenomena developing in the melt, such as carbon-concentration distribution and melt flow, cannot be observed in situ. Thirdly, the fluid flow developing in the melt is very complex since it is driven by competing forces such as the electromagnetic field of the induction-heating coils, buoyancy, and the surface tension gradient along the free surface. Thus, numerical simulation has been considered as a feasible and inexpensive tool for studying the transport phenomena in the melt of the TSSG method.
In this direction, Yamamoto et al. (2017) considered numerically a higher growth temperature than the conventional one in order to increase carbon solubility to achieve a higher crystal growth rate. In their study, it was predicted a strong Marangoni flow in the melt that becomes downward in the region under the seed crystal. This downward flow under the seed leads to a non-uniform crystal growth rate. Wang et al. (2018) investigated the effect of the static magnetic field in the TSSG growth of SiC. It was found that the application of a cusp magnetic field together with seed rotation was beneficial to suppress such a Marangoni flow.

The optimization of the associated control parameters of the melt could not have been done in their study since such optimization required high computational costs due to the complex nature of the fluid flow involved. Then, Takehara et al. (2020) introduced the Bayesian optimization algorithm and efficiently optimized the control parameters to suppress the Marangoni convection. From the comparison of the melt flows before and after the optimization, they concluded that the upward flow below the seed crystal in the optimal state is the key to improve the crystal growth rate.

However, in such a classical approach for comparison, the determination of which region in the melt is significant in terms of the effect of the melt flow highly depends on the user’s interpretation and observation. Such a subjective prediction can be overcome by the introduction of an “automatic recognition technique”. To this end, in this study, we developed a new strategy to determine the sensitive (key) region in the melt flow using sensitivity analysis of a convolutional neural network (CNN) that predicts the objective function (crystal growth rate) from the input (melt flow). This auto-recognition technique does not rely on the observation that the Marangoni convection is an important parameter affecting the crystal growth rate, but it successfully determines the characteristic flow pattern in the optimized melt flow.

The results of the Bayesian optimization were updated for a more optimal state, and then the optimal state was analyzed by both the classical approach such as given by Takehara et al. (2020) and also by the “automatic recognition technique”.

2. Numerical methodology

2.1. Numerical simulation of the TSSG method

The simulation of the TSSG method (Fig. 1) was carried out using Integrate Process Model (IPM) by Derby et al. (1989); Derby and Gresho (1987) separately for (i) the electromagnetic field, (ii) the temperature field in the whole furnace, and (iii) the transport phenomena in the melt as same as the previous studies by Yamamoto et al. (2017); Wang et al. (2019); Horiuchi et al. (2019); Takehara et al. (2020). The static cusp magnetic field, \( B \), is applied to the melt as in Wang et al. (2018):

\[
\frac{B}{B_0} = \left( r, 1, 2(H_0 - z) \right)/R_c
\]

where \( B_0 \) is the intensity of the cusp magnetic field, \( H_0 \) the axial coordinate of the center of the cusp magnetic field, and \( R_c \) radius of the crucible (see Fig. 1). The origin of the cylindrical coordinate is at the center of the crucible bottom surface. In the present simulation model, we assume that the growth melt is an incompressible and Newtonian fluid, the physical properties of Si can be used for the melt since the solubility of carbon is low, the melt physical properties can be taken as constant, and the deformation of the free surface due to the melt flow and the meniscus effect is negligible. Using the Boussinesq approximation, the dimensionless governing equations

![Fig. 1 The schematic cross-sectional view of the TSSG system (left), and the graph showing the cusp magnetic field (blue glyph) and the seed rotation (orange glyph) (right).](image)
of the melt are
\[
\frac{\partial u^*}{\partial t^*} + (u^* \cdot \nabla) u^* = -\nabla^* p^* + \frac{1}{Re} \nabla^2 u^* + \frac{1}{Re} \left( \frac{H_{\text{coil}}}{\nu} f_E + \frac{H_{\text{cusp}}}{\nu} j^* \times B^* \right) - \frac{Gr}{Re^2} \frac{g}{|g|} T^* \tag{1}
\]
\[
\nabla^* \cdot u^* = 0 \tag{2}
\]
\[
\frac{\partial T^*}{\partial t^*} + (u^* \cdot \nabla) T^* = \frac{1}{Re Pr} \nabla^2 T^* + \frac{Gr^3}{Re} \frac{Ec}{Pr} S^* \tag{3}
\]
\[
\frac{\partial \bar{c}^*}{\partial t^*} + (u^* \cdot \nabla) \bar{c}^* = \frac{1}{Sc Re} \nabla^2 \bar{c}^* \tag{4}
\]
where \(u = (u_x, u_y, u_z)\) is the melt flow velocity, \(t\) time, \(\rho\) the melt density, \(p\) the melt pressure, \(f_E\) the Lorentz force density, \(j^*\) the induced electric current density, \(B\) the cusp magnetic field, \(g\) the gravitational acceleration, \(T\) the melt temperature, \(S\) the Joule heat generation density, \(\bar{c}\) the carbon concentration and \((\cdot)^*\) represents dimensionless variables.

The dimensionless numbers are
\[
Re = \frac{U_{\text{ref}} R_c}{\nu}, \quad H_{\text{coil}} = \frac{\sigma_s}{\nu p B_{\text{ref}} R_c}, \quad H_{\text{cusp}} = \frac{\sigma_s}{\nu p B_{\text{ref}} R_c}, \quad Gr = \frac{\beta \rho R_c^3 (T_{\text{max}} - T_{\text{min}})}{\nu^2}, \quad Pr = \frac{v}{\alpha}, \quad Ec = \frac{U_{\text{ref}}^2}{C_p (T_{\text{max}} - T_{\text{min}})}, \quad Sc = \frac{v}{D}, \quad Ma = \frac{\partial \bar{r}}{\partial T} \rho p (T_{\text{max}} - T_{\text{min}}) \frac{1}{\nu}\tag{5}
\]
where \(U_{\text{ref}}\) is the reference velocity, \(\sigma_s\) the melt electrical conductivity, \(B_{\text{ref}}\) the reference magnetic field of induction coils, \(\beta\) the melt thermal expansion coefficient, \(T_{\text{max}}\) the maximum temperature of the melt, \(T_{\text{min}}\) the minimum temperature of the melt, \(\alpha\) the melt thermal diffusivity, \(C_p\) the specific heat, and \(D\) the diffusion coefficient of \(\bar{c}\) in the melt.

The no-slip boundary condition for the flow velocity is applied on the crucible wall and the seed surface. The Marangoni boundary condition is used along the free surface expressed as
\[
\frac{\partial u_x^*}{\partial z^*} = \frac{Ma}{Pr Re} \frac{\partial T^*}{\partial z^*}. \tag{7}
\]

For the silicon concentration in the melt, the equilibrium concentration boundary condition was used on the seed and along the crucible wall as in Durand and Duby (1999). The physical properties used in the simulation for the TSSG system were taken from Yamamoto et al. (2017); Durand and Duby (2010) The governing equations of the melt are solved by using the OpenFOAM. As the initial condition of the simulation, a fully developed flow subjected to an applied cusp magnetic field \(B\) is used. After changing the control parameters, the transition period of the initial 50 seconds (15 dimensionless times) is discarded.

The crystal growth rate \(v_g\) was calculated as the mean value of 100 seconds (30 dimensionless times) after the initial 50 seconds, and given by
\[
v_g = -\frac{D \bar{M}_{\text{SiC}}}{\rho_{\text{SiC}}} \frac{\partial \bar{c}}{\partial z}|_{\text{seed}} \tag{8}
\]
where \(\bar{M}_{\text{SiC}}\) and \(\rho_{\text{SiC}}\) are the molar weight and the mass density of SiC crystal respectively. The objective function is defined as
\[
f = M^* - \sigma^*. \tag{9}
\]
using the normalized mean \(M^*\) and deviation \(\sigma^*\) of the growth rate profile, \(v_g\), as in the previous study (Takehara et al., 2020).

\[
M^* = M_{g*}/M_0, \quad \sigma^* = \sigma_{g*}/\sigma_0 \tag{10}
\]
where
\[
M_{g*} = \frac{1}{L} \int_0^L v_g(r) \, dr, \quad \sigma_{g*} = \sqrt{\frac{1}{L} \int_0^L [v_g(r) - M_{g*}]^2 \, dr} \tag{11}
\]
\[
M_0 = M_{g*}|_{B=\Omega=0}, \quad \sigma_0 = \sigma_{g*}|_{B=\Omega=0} \tag{12}
\]
and \(L\) is the length of the target region (\(L = 4\) mm).
2.2. The Bayesian optimization

The three control parameters, \( x = (B_0, H_0, \Omega) \), were considered and their discrete values in a realistic parameter range were chosen for the optimization as shown in Table 1. The Bayesian optimization maximizes an objective function \( y = f(x) \), where \( x \) is a parameter vector and \( y \) is an output value. The optimization algorithm (see also Fig. 2) is as follows.

1. Making the initial dataset, \( D_0 = \{ (x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(N)}, y^{(N)}) \} \).
2. Gaussian process regression to calculate the predictive mean function, \( \mu_{\text{GPR}}(x^{N+1}) \), and the standard deviation functions, \( \sigma_{\text{GPR}}(x^{N+1}) \) (see Appendix B).
3. Calculation with the next control parameter vector using an acquisition function:
   \[ x^{N+1} = \arg \max_x \{ a(\mu_{\text{GPR}}, \sigma_{\text{GPR}}) \} \]
4. CFD simulation with the updated parameter \( x^{N+1} \) and the objective function \( f = y^{(N+1)} \) is calculated.
5. Renewing the dataset \( D_{N+1} \) by adding \( (x^{(N+1)}, y^{(N+1)}) \), and go to the step (2).

At first, the optimization was performed using discrete points in the parameter space shown in Table 1 as in Takehara et al. (2020) and, then further optimization was performed again using continuous values of the parameters. The former and latter optimization (for discrete and continuous parameters) are called discrete-Bayesian optimization and continuous-Bayesian optimization in this study. The Bayesian optimization is conducted by using the GPyOpt library.

2.3. Explainable machine learning for the optimized state

As previous studies have predicted, the Marangoni convection in the TSSG melt affects the SiC crystal growth rate, and the upward flow developing below the seed crystal leads to high- and uniform-crystal growth rates (Horiuchi et al. (2019); Takehara et al. (2020)). However, the analysis of the optimized melt flow and the determination of which region has significant effects highly depends on the observer’s experience and interpretation. As mentioned earlier, such an approach needs to be replaced by an automatic recognition technique. In this study, a new strategy is introduced to determine the sensitive region in the melt flow to the objective function and also to explain the reason for the optimal state being automatically determined only by data. To this end, first, a convolutional neural network (CNN) that predicts the objective function \( f \) from the mean flow velocity field in the melt is constructed. Then, the part of the convective flow that contributes to the prediction of CNN is visualized using a sensitivity map. The objective function predicted by CNN from the flow data \( U = (u_r, u_\theta, u_z) \) is defined as \( f_{\text{CNN}}(U) \). A sensitivity map \( S(U) = (S_r, S_\theta, S_z) \) is defined as

\[
S(U) = \frac{\partial (f_{\text{CNN}})}{\partial U}
\]

which represents the sensitivity of \( f_{\text{CNN}} \) with respect to the deviation of \( U \). When the sensitivity is large, the convection is important to predict \( f_{\text{CNN}} \). However, it is difficult to consider the meaning of the sensitivity map since the map is not robust due to the noise in data. Smilkov et al. (2017) suggested a method to remove the noise adding Gaussian noise to the input data, \( U \). This method is called SmoothGrad. The sensitivity map \( \hat{S}(U) \) using this technique is described as

\[
\hat{S}(U) = \frac{1}{n} \sum_{i=1}^{n} S(U + N(0, \sigma^2))
\]
Table 1 The discrete values of the control parameters: $B_0$, the intensity of the cusp magnetic field; $H_0$, the position of the cusp magnetic field; $\Omega$, the seed rotation speed (left), and the corresponding non-dimensional numbers (right).

| $B_0$ [T] | $H_0$ [10^{-3}m] | $\Omega$ [s^{-1}] | $H_0$ | $H_0/R_c$ | $Re_{\Omega}$ |
|-----------|------------------|------------------|--------|-----------|--------------|
| 0         | 0                | 0                | 0      | 0         | 0            |
| 0.05      | 5                | 0.33             | 43.6   | 0.22      | 750          |
| 0.1       | 10               | 0.67             | 87.2   | 0.44      | 1500         |
| 0.15      | 15               | 1.00             | 130.8  | 0.67      | 2250         |
| 0.2       | 20               | 1.33             | 174.4  | 0.89      | 3000         |

where $n$ is the number of noise-augmented samples ($n = 100$), and $\mathcal{N}(0, \sigma^2)$ presents the Gaussian noise with standard deviation $\sigma$ ($\sigma = 0.1$). $S$ and $\hat{S}$ have the same dimensions as $U$. The magnitude of the sensitivity map $|\hat{S}|$ is defined as

$$|\hat{S}(U)| = \sqrt{\hat{S}^2 + \hat{S}_\theta^2 + \hat{S}_\phi^2},$$

(15)

3. Results and discussion

Takehara et al. (2020) stopped the discrete-Bayesian optimization algorithm at the 24th iteration since the objective function, $f$, could not have been updated for more than 10 iterations. The optimal parameters were obtained as ($B_0$, $H_0$, $\Omega$) = (0.05 T, 5 x10^{-3}m, 1.33 s^{-1}) and the corresponding $H_0$ = 43.6, $H_0/R_c$ = 0.22, $Re_\Omega = \Omega R_c \omega/\nu = 3000$. After the discrete optimization, in the present study, the continuous Bayesian optimization was conducted at 5 iterations and the optimal state was updated. The optimal parameters were renewed as ($B_0$, $H_0$, $\Omega$) = (0.046 T, 3 x10^{-3}m, 1.33 s^{-1}) and the corresponding $H_0$ = 40.1, $H_0/R_c$ = 0.132, $Re_\Omega = 3000$. The map ($M^*$, $\sigma^*$) of the results of the discrete- and continuous-Bayesian optimization from the initial 8 calculations (with $\times$) is shown in Fig 3(a). The upper left and lower right means the higher and lower value of $f = (M^* - \sigma^*)$.

The optimal value was efficiently searched during the continuous Bayesian optimization represented as the star dots in Fig. 3(a) and the Pareto line was also obtained successfully. The crystal growth rates of the initial and the optimal states are presented in Fig 3(b). A high- and uniform-growth rate is successfully obtained in the target region.

3.1. Classical explanation approach for optimal flow

Figure 4 shows the flow velocity, temperature, and concentration fields in the initial (a,c,e) and optimal (b,d,f) states. It is found that the downward flow below the seed crystal in the initial state becomes the upward flow in the optimal state as reported in Takehara et al. (2020). The downward flow arises due to the effect of Marangoni convection (Yamamoto et al., 2017), and the Marangoni convection is suppressed in the optimal state. The transport of carbon to the seed crystal in the optimal state are more uniform than the initial state because of the upward flow in the melt below the seed. The uniform transport of carbon leads to a high- and uniform-crystal growth rate.

We analyze the contributions of each driving force term in Eq. (1), i.e., the Lorentz forces ($F_E$), the gravity body force, and the Marangoni effect under the applied cusp magnetic field. Figure 5 shows the flow patterns of the electromagnetic flow in (a), the Marangoni convection (b), and the buoyancy flow (c) under the optimal cusp magnetic field ($H_0$ = 40.1, $H_0/R_c$ = 0.132) and the seed rotation ($Re_\Omega = 3000$). It is found that the effect of buoyancy is not significant as seen in Fig. 5(c), and its effect is then considered negligible. By comparing the flow pattern of optimal state (Fig. 4(d)) and the electromagnetic flow (Fig. 5(a)), it is seen that the effect of the electromagnetic field of the induction-heating coils is dominant in the melt because both flow structures are similar. Figure 5(a) also shows that the electromagnetic force leads to an upward flow in the optimal state below the seed. Figure 5(b) shows that the applied cusp magnetic field bends the Marangoni downward flow towards outside. As a consequence, the effect of Marangoni convection under the seed decreases, and the upward flow by the electromagnetic force arises in the optimal state. The vortex predicted near the edge of the seed in Fig. 4(b) is generated by the interaction of the electromagnetic flow (Fig. 5(a)) with the Marangoni convection from the liquid-air surface (Fig. 5(b)). The vortex develops because the Marangoni flow separates from the seed surface near the edge due to high-pressure in this region induced by the upward flow induced by the electromagnetic body force. Figure 6 shows the $\partial u_z/\partial z$ along the seed surface (a), the concentration gradient $\partial c/\partial z$ (b), and the concentration fields with streamlines (c) near the seed edge. The flow separation drastically changes the concentration boundary layer thickness, and a smaller crystal-growth rate is predicted near the seed edge. The peak
Fig. 3 (a) The distribution of $M^*$ and $\sigma^*$ where the gray dots represent the whole parameter combination (125 calculations) and the colored circles are the 24 calculations by the discrete-Bayesian optimization (Takehara et al., 2020). The star symbols indicate five more iterations from the discrete-optimal parameters using the continuous-Bayesian optimization algorithm. Some of the star symbols are on a Pareto line. The circles with cross indicate the initial dataset for the Bayesian optimization. (b) The crystal growth rate, $v_g$, in the initial and the optimal states.

Fig. 4 The velocity and temperature field below the seed (a,b), and in the melt (c,d) of the initial and optimal state. Concentration field and streamlines (e,f): (a,c,e) $Ha_{cusp} = 0$, (b,d,f) the optimal state. The optimal state: the continuous optimal state.

position of A in Fig. 3(b) corresponds roughly to where the separation point is ($\partial u_r/\partial z = 0$ in Fig. 6(a)). And, as in the position of B in Fig. 6(b), the concentration boundary layer becomes slightly thinner and the concentration gradient becomes higher due to the peak of $\partial u_r/\partial z$ in Fig. 6(a), which leads to the local maximum of crystal-growth rate probably due to the high-$Sc$ number effect ($Sc = 18.5$).

3.2. Explanation using a machine-learning technique

The mean absolute error (MAE) of the training and test data (velocity fields) for the present CNN (see Sec. 4.1) is 0.016 and 0.096, respectively. After the construction of CNN, SmoothGrad is applied to the CNN in order to know where is sensitive to the objective function $f$ in Eq. (9). Figure 7 shows the sensitivity maps (Eq. (15)) of the initial and the optimal velocity fields. The sensitivity maps of both states have similar patterns near the seed crystal edge. Since the effect of the Marangoni convection is significantly large near the seed edge (A), the sensitivity maps successfully indicate that the Marangoni convection plays an important role to the objective function $f$ as suggested by the previous studies (Yamamoto et al., 2017; Horiuchi et al., 2019). In the initial state, the sensitivity is also large around the seed center (region C) as in the region near the seed edge (region A) and where the convection is seen along the central vertical axis ($r = 0$) (region B). In the optimal state, however, the sensitivity is much smaller in the regions around the central vertical axis (region B) and near the seed center (region C). This is because the convections in regions B and C are sufficiently improved not to affect $f$ in the optimal state. This result indicates that the upward flow predicted in Regions B and C is a key factor for the optimal state, which agrees with the explanation done by the classical approach. Therefore, we can conclude that the convection pattern is improved in the regions where we have high sensitivity in the initial state and low sensitivity in the optimal state.

Moreover Fig. 8(a–f) shows the same sensitivity maps for the velocity components: $u_r$, $u_\theta$, and $u_z$. The contribution of $u_z$ component is dominant both in the initial and optimal states. Figure 8(c) shows a positive sensitivity around the seed
Fig. 5  The velocity vectors representing the contributions of each driving forcing term under the optimal cusp magnetic field ($H_{0cusp} = 40.1$, $H_0/R_c = 0.132$) and the seed rotation ($Re_Ω = 3000$): (a) $Ma = 0$, $Gr = 0$, $|F^*_L| = 9700$; (b) $Ma = 1.0 \times 10^4$, $Gr = 0$, $|F^*_L| = 0$; (c) $Ma = 0$, $Gr = 6 \times 10^6$, $|F^*_L| = 0$.

Fig. 6  (a) $\partial u_r/\partial z$ along the seed surface near the seed edge. $\partial u_r/\partial z = 0$ indicates the separation of the velocity boundary layer. (b) $-\partial \bar{c}/\partial z$ distribution near the seed edge. (c) The concentration fields with the streamlines near the seed edge. The points of A and B in (b) correspond to those in Fig. 3(b).

The positive sensitivity indicates that the upward flow toward the seed center increases $f$ which is realized in the optimal state. These predictions agree with those of the previous numerical optimization of TSSG. Horiuchi et al. (2019) reported that the development of an upward flow under the seed is necessary for a high- and uniform-crystal growth rate. It is also reported that, when the Marangoni convection along the free surface of the melt is sufficiently suppressed, the seed-edge vortex appears as a separation bubble as shown in Fig. 4(b). The sensitivity analysis of the present CNN model also detects such regions as shown in Fig. 8(c, f). However, Fig. 8(f) still shows a negative sensitivity near the seed edge even in the optimal state. There may be a more favorable convection field that cannot be achieved by the present cusp field or by seed rotation. The negative sensitivity in Region A indicates that a stronger downward flow by the seed edge vortex, driven by the separation of the Marangoni convection, may further increase $f$, which has not been clearly detected in previous studies by the classical approach. The sensitivity analysis using the present CNN shows the possibility of further improvements in the quality of SiC crystals grown by the TSSG technique.

4. Conclusions

The Bayesian optimization was applied to the TSSG method of SiC under the effects of static cusp magnetic field and seed rotation. From the present study, the more optimal values of the magnetic field intensity $B_0$, the position of the cusp magnetic field $H_0$, and the seed rotation speed $Ω$ were obtained as $B_0 = 0.046$ T, $H_0 = 3 \times 10^{-3}$ m, $Ω = 1.33$ s$^{-1}$ (and the corresponding non-dimensional number $H_{0cusp} = 40.1$, $H_0/R_c = 0.132$, $Re_Ω = QR_Ω/R_{seed}/ν = 3000$), respectively.

By comparing the initial and optimal states, it was found that the Marangoni flow bends toward the crucible wall, and its effect below the seed weakens in the optimal state. As a result, an upward flow toward the seed develops. This flow is responsible for the improvement in the crystal growth rate. The change in the concentration boundary layer thickness due to flow separation leads to a non-uniform crystal growth rate near the seed edge in the optimal state.

A new proposed data-driven sensitivity analysis using CNN and SmoothGrad automatically predicts that the upward flow below the seed crystals is the key factor of the objective function $f$, which is in agreement with the prediction made by the classical comparison approach. Furthermore, the present sensitivity analysis also predicts for the first time that the
Fig. 7 The sensitivity maps of all velocity fields. The red and blue region represents high and low sensitivity to $f$. (Left) the initial state. (Right) the optimized state obtained by the Bayesian algorithm. The arrows represent the corresponding velocity vectors ($u_r, u_z$). The region A, B and C are the high sensitivity regions where we focus (see in the text).

Fig. 8 The sensitivity map for each velocity component, (a,d) $u_r$, (b,e) $u_\theta$, (c,f) $u_z$. (a-c) the initial state, (d-f) the optimal states obtained by the Bayesian optimization. The color shows the sensitivity $\hat{S}$.

$u_z$ flow velocity component at the seed edge enhances the objective function $f$.

This method tells us where the important region would be, which is known by the constructed CNN implicitly. Such an auto-recognition of physical insights for optimization will be very valuable when we consider more complex phenomena and encounter a huge data set in future studies.

Acknowledgments

This research uses computational resources under Research Institute for Information Technology, Kyushu University. The research work was financially supported by Grant-in-Aid for Scientific Research (A) (JSPS KAKENHI Grant Number JP18H03839) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Appendix A: Gaussian process

The parameter vector is expressed as $x = [x_1, x_2, \cdots, x_m]^T$ where $m$ is the number of the parameter. When the feature vector of $x$ is expressed as $\phi(x) = [\phi_0(x), \phi_1(x), \cdots, \phi_H(x)]^T$, a linear regression model is obtained as

$$y = \sum_{i=0}^H w_i \phi_i(x) = w^T \phi(x)$$

(A.1)
where \( w = [w_0, w_1, \ldots, w_M]^T \) is the weights of each feature vector. Regarding the initial dataset, Eq. (A.1) is expressed as

\[
\begin{pmatrix}
y^{(1)}_i \\
y^{(2)}_i \\
\vdots \\
y^{(N)}_i
\end{pmatrix} = \begin{pmatrix}
\phi_0(x^{(1)}_i) & \phi_1(x^{(1)}_i) & \cdots & \phi_H(x^{(1)}_i) \\
\phi_0(x^{(2)}_i) & \phi_1(x^{(2)}_i) & \cdots & \phi_H(x^{(2)}_i) \\
\vdots & \vdots & \ddots & \vdots \\
\phi_0(x^{(N)}_i) & \phi_1(x^{(N)}_i) & \cdots & \phi_H(x^{(N)}_i)
\end{pmatrix} \begin{pmatrix}
w_0 \\
w_1 \\
\vdots \\
w_M
\end{pmatrix} = \Phi w
\]  

(A.2)

It is assumed that \( w \) follows Gaussian distribution with mean 0 and variance \( \lambda^2 I \)

\[
w \sim \mathcal{N}(0, \lambda^2 I)
\]  

(A.3)

where \( I \) is an identity matrix. The assumption means \( w \) is independent of each other. From this assumption, \( y \) also follows Gaussian distribution, and the expected value \( E[y] \) and covariance matrix \( K \) of the Gaussian distribution of \( y \) are as follows.

\[
E[y] = E[\Phi w] = \Phi E[w] = 0
\]  

(A.4)

\[
K = E[y y^T] - E[y]E[y]^T = E[(\Phi w)(\Phi w)^T] = \Phi E[w w^T] \Phi^T = \lambda^2 \Phi \Phi^T
\]

That is, \( y \) follows multivariate Gaussian distribution.

\[
y \sim \mathcal{N}(0, \lambda^2 \Phi \Phi^T)
\]  

(A.6)

When Eq. (A.6) is true with any input vector \( x \), the relationship between \( x \) and \( y \) follows the Gaussian process. Equation (A.6) means \( y \) distribution is obtained only by \( \lambda^2 \Phi \Phi^T \) without \( w \).

The \((i, j)\)-th element \( K_{i,j} = \lambda^2 \Phi \Phi^T_{i,j} \) is written as

\[
K_{i,j} = \lambda^2 \Phi \Phi^T_{i,j} = \lambda^2 \sum_{k=0}^M \phi_k(x^{(i)}) \phi_k(x^{(j)}) = \lambda^2 \phi(x^{(i)})^T \phi(x^{(j)})
\]  

(A.7)

If the value of \( K_{i,j} \) is obtained in any way, it is unnecessary to determine the feature vector \( \phi \). In the Gaussian process, a kernel function \( k(x^{(i)}, x^{(j)}) \) is used in order to calculate \( K_{i,j} \).

\[
k(x^{(i)}, x^{(j)}) = K_{i,j} = \lambda^2 \phi(x^{(i)})^T \phi(x^{(j)})
\]  

(A.8)

In this study, Matérn kernel 5/2 was used:

\[
k(x^{(i)}, x^{(j)}) = \theta_0 \left( 1 + \frac{\sqrt{5l}}{\theta_1} + \frac{5l^2}{3\theta_1^2} \right) \exp \left( -\frac{\sqrt{5l}}{\theta_1} \right)
\]  

(A.9)

\[
l = |x^{(i)} - x^{(j)}|
\]  

(A.10)

where \( \theta_0, \theta_1 \) are hyperparameters optimized by maximum likelihood estimation of \( \log p(y|x, \theta) \).

**Appendix B: Gaussian process regression**

The output vector \( y = [y^{(1)}, y^{(2)}, \ldots, y^{(N)}]^T \) of the initial dataset \( D \) follows Gaussian distribution (Eq. (A.6)). About an unknown \( x^{(N+1)}, y^{(N+1)} \), \( y^* = [y^{(1)}, y^{(2)}, \ldots, y^{(N)}, y^{(N+1)}]^T \) also follows Gaussian distribution.

\[
y^* \sim \mathcal{N}(0, K^*)
\]  

(A.11)

And kernel matrix \( K^* \) is

\[
K^* = \begin{pmatrix}
k(x^{(1)}, x^{(1)}) & k(x^{(1)}, x^{(2)}) & \cdots & k(x^{(1)}, x^{(N)}) & k(x^{(1)}, x^{(N+1)}) \\
k(x^{(2)}, x^{(1)}) & k(x^{(2)}, x^{(2)}) & \cdots & k(x^{(2)}, x^{(N)}) & k(x^{(2)}, x^{(N+1)}) \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
k(x^{(N)}, x^{(1)}) & k(x^{(N)}, x^{(2)}) & \cdots & k(x^{(N)}, x^{(N)}) & k(x^{(N)}, x^{(N+1)}) \\
k(x^{(N+1)}, x^{(1)}) & k(x^{(N+1)}, x^{(2)}) & \cdots & k(x^{(N+1)}, x^{(N)}) & k(x^{(N+1)}, x^{(N+1)})
\end{pmatrix}
\]  

(A.12)
where $k^* = [k(x^{(1)}, x^{(N+1)}), k(x^{(2)}, x^{(N+1)}), \ldots, k(x^{(N)}, x^{(N+1)})]^T$. Generally, when a vector following Gaussian distribution satisfies
\[
\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right),
\]
the conditional probability $p(y_2 | y_1)$ that $y_2$ is obtained after $y_1$ is obtained is expressed as
\[
p(y_2 | y_1) \sim \mathcal{N} \left( \mu_2 + \Sigma_{21}(y_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \right).
\]
Comparing Eq. (A.11) and Eq. (A.14), the conditional probability $p(y^{(N+1)} | y)$ is expressed as
\[
p(y^{(N+1)} | y) \sim \mathcal{N} \left( \mu_{GPR}(x^{(N+1)}), \sigma_{GPR}(x^{(N+1)}) \right).
\]
\[
\mu_{GPR}(x^{(N+1)}) = k^{\top} K^{-1} y
\]
\[
\sigma_{GPR}(x^{(N+1)}) = k(x^{(N+1)}, x^{(N+1)}) - k^{\top} K^{-1} k^*.
\]

\[\text{Fig. 9} \quad \text{An example of the Gaussian process regression for } x \text{ when } N = 5. \text{ The vertical dashed line is the current dataset, } D. \text{ The gray area is the confidence band of the bounds of } \pm 1.96 \sigma_{GPR}. \text{ The red line is the Gaussian distribution at the next parameter } x^{(N+1)}.\]

\section*{Appendix C: Acquisition functions}

From the Gaussian process regression, the next control parameter vector $x_{\text{next}} = x^{(N+1)}$ is obtained using an acquisition function. Some known acquisition functions are the following:
\[
a_{\text{UCB}}(x^{(N+1)}) = \mu_{GPR}(x^{(N+1)}) + s \times \sigma_{GPR}(x^{(N+1)})
\]
\[
a_{\text{PI}}(x^{(N+1)}) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} |\sigma_{GPR}(x^{(N+1)})|^2} \exp \left\{ -\frac{1}{2 |\sigma_{GPR}(x^{(N+1)})|^2} \left( y - \mu_{GPR}(x^{(N+1)}) \right)^2 \right\} dy
\]
\[
a_{\text{EI}}(x^{(N+1)}) = \left( \mu_{GPR}(x^{(N+1)}) - y_{\text{max}} \right) a_{\text{PI}} + \left| \sigma_{GPR}(x^{(N+1)}) \right|^2 \frac{1}{\sqrt{2\pi} |\sigma_{GPR}(x^{(N+1)})|^2} \exp \left\{ -\frac{1}{2 |\sigma_{GPR}(x^{(N+1)})|^2} \left( y_{\text{max}} - \mu_{GPR}(x^{(N+1)}) \right)^2 \right\}
\]

$a_{\text{UCB}}$ means that confidence interval. When $s = 1.96$, $a_{\text{UCB}}$ expresses a 95% confidence interval. $a_{\text{PI}}$ means that probability to improve the maximum value $y_{\text{max}}$ and $a_{\text{EI}}$ expected value to improve the maximum value. In this study, $a_{\text{EI}}$ is used. The next parameter $x_{\text{next}}$ is calculated by $x_{\text{next}} = \arg\max_x a(x)$.
Appendix D: Convolutional neural network

A convolutional neural network (CNN) is frequently used for image recognition and CNN handles two or three-dimensional data, such as gray-scale pictures (vertical pixel × horizontal pixel) or color pictures (vertical pixel × horizontal pixel × color channels of RGB) (Krizhevsky et al., 2012). In the present study, the time-averaged velocity fields are handled as three-dimensional-color pictures (vertical grids × horizontal grids × velocity components \((u_r, u_\theta, u_z)\)), and CNN predicts the objective function \(f\) from the calculated convection field data. The model has four convolution layers with a max pooling layer, and two fully connected layers with 16 neurons. Zero padding is done in the first and second convolution layers as shown in Fig. 10. The detailed information of each layer is in Table 2. Rectified linear unit (ReLU), i.e. a ramp function, is used in all layers as an activation function.

The present CNN was trained using 125 data of full combinations of the parameters in Table 1, 5 data obtained from the continuous-Bayesian optimization, and 16 data from random sampling. The accuracy of the model was confirmed using 25 test data. In training the present CNN, each component of the input data \(U\) is filtered from the full-size data \((480 \times 450)\) to \((300 \times 300)\) grids data to reduce the calculation costs of CNN and the only \(f\) is scaled by its mean and standard deviation.

Table 2 The layer information of the CNN model in the present study.

| Layer               | Filter size | Stride | Output size       |
|---------------------|-------------|--------|-------------------|
| input               |             |        | \((300, 300, 3)\) |
| Convolution 1       | (3, 3)      | (1, 1) | \((300, 300, 16)\)|
| Max pooling 1       | (2, 2)      | (2, 2) | \((150, 150, 16)\)|
| Convolution 2       | (3, 3)      | (1, 1) | \((150, 150, 32)\)|
| Max pooling 2       | (2, 2)      | (2, 2) | \((75, 75, 32)\)  |
| Convolution 3       | (3, 3)      | (1, 1) | \((75, 75, 64)\)  |
| Max pooling 3       | (2, 2)      | (2, 2) | \((36, 36, 64)\)  |
| Convolution 4       | (3, 3)      | (1, 1) | \((34, 34, 64)\)  |
| Fully connected 1   |             |        | 16                |
| Fully connected 2   |             |        | 1                 |

Nomenclature

\(B_0\) cusp magnetic field (T)
\(B_{EM}\) magnetic field of induction-heating coil (T)
\(\bar{c}\) concentration (mol/m³)
\(a\) acquisition function (-)
\(C_p\) specific heat (J/(kg·K))
\(B\) magnetic field (T)
\(D\) diffusion coefficient (m²/s)
$\mathcal{D}_N$ dataset with $N$ elements (-)

$Ec$ Eckert number (-)

$f$ objective function (-)

$f_{\text{CNN}}$ objective function value predicted by CNN (-)

$F_E$ Lorentz force density (N/m$^3$)

$g$ gravity acceleration (m/s$^2$)

$Gr$ Grashof number (-)

$H$ the number of feature vector (-)

$H_0$ axial coordinate of the center of the cusp magnetic field (m)

$H_{ai}$ Hartman number of $i$ (-)

$I$ identity matrix (-)

$j$ current density (A/m$^2$)

$k$ kernel function (-)

$K$ covariance matrix of Gaussian distribution (-)

$L$ target region of optimization (m)

$m$ the number of parameters (-)

$M^*$ dimensionless mean value of crystal growth rate (-)

$M_{gr}$ mean value of crystal growth rate ($\mu$m/h)

$Ma$ Marangoni number (-)

$\bar{M}_{\text{SiC}}$ molar weight of SiC (kg/mol)

$p$ pressure (Pa)

$Pr$ Prandtl number (-)

$r$ radial coordinate (m)

$R_c$ radius of crucible (m)

$Re$ Reynolds number (-)

$S$ sensitivity using raw gradients (s/m)

$Sc$ Schmidt number (-)

$\dot{S}$ Joule heat generation density (W/m$^3$)

$\hat{S}$ sensitivity by smoothGrad (s/m)

$t$ time (s)

$T$ temperature (K)

$T_{\text{max}}$ maximum temperature (K)

$T_{\text{min}}$ minimum temperature (K)

$u$ fluid velocity (m/s)

$U_{\text{ref}}$ reference velocity (m/s)

$v_g$ crystal growth rate ($\mu$m/h)

$w$ weight of feature vector (-)

$y$ output vector (-)

$x$ control-parameter vector (-)

$x_{\text{next}}$ next parameter vector of Bayesian optimization (-)

$y_{\text{max}}$ the maximum value of $y$ (-)

$z$ axial coordinate (m)

**Greek letters**

$\alpha$ thermal diffusivity (m$^2$/s)

$\beta$ thermal expansion coefficient (1/K)

$\Omega$ rotation speed of seed crystal (rpm)

$\mu_{\text{GPR}}$ predictive mean function of Gaussian process regression (-)

$\nu$ kinematic viscosity (m$^2$/s)

$\phi$ feature vector (-)

$\rho$ density (kg/m$^3$)

$\rho_{\text{SiC}}$ density of SiC (kg/m$^3$)

$\sigma$ surface tension (N/m)

$\sigma^*$ dimensionless standard deviation value of crystal growth rate (-)

$\sigma_0$ standard deviation value of crystal growth rate (-)

$\sigma_e$ electrical conductivity (S/m)

$\sigma_{\text{GPR}}$ standard deviation function of Gaussian process regression (-)

$\theta$ radial angular (-)

$\theta_0$, $\theta_1$ hyperparameter of a kernel function (-)
References

Derby JJ., Gresho PM., A finite element model for induction heating of a metal crucible, Journal of crystal growth, Vol.85, No.1-2 (1987), pp.40-48.
Derby JJ., Atherton LJ., Gresho PM., An integrated process model for the growth of oxide crystals by the czochralski method, Journal of crystal growth, Vol.97, No.3-4 (1989), pp.792–826.
Durand F., Duby JC., Carbon solubility in solid and liquid silicon – A review with reference to eutectic equilibrium, Journal of Phase Equilibria, Vol.20, No.1 (1999), pp.61–63.
Durand F., Duby JC., Coupled heat transfer and fluid dynamics modeling of high-temperature SiC solution growth. Journal of Crystal Growth, Vol.312, No.2 (2010), pp.155–163.
Horiuchi T., Wang L., Sekimoto A., Okano Y., Ujihara T., Dost S., Adjoint-based sensitivity analysis for the optimal crucible temperature profile in the RF-Heating TSSG-SiC crystal growth process, Journal of Crystal Growth, Vol.517 (2019), pp.59–63.
Krizhevsky A., Sutskever I., Hinton GE., Imagenet classification with deep convolutional neural networks, Advances in neural information processing systems 25 (NIPS 2012) (2012), pp.1097–1105.
Smilkov D., Thorat N., Kim B., Viégas F., Wattenberg M., Smoothgrad: removing noise by adding noise. JarXiv preprint arXiv:170603825 (2017).
Takehara Y., Sekimoto A., Okano Y., Ujihara T., Dost S., Bayesian optimization for a high-and uniform-crystal growth rate in the top-seeded solution growth process of silicon carbide under applied magnetic field and seed rotation, Journal of Crystal Growth, Vol.532, No.125437, (2020).
Wang L., Horiuchi T., Sekimoto A., Okano Y., Ujihara T., Dost S., Numerical investigation of the effect of static magnetic field on the TSSG growth of SiC. Journal of Crystal Growth, Vol.498 (2018), pp.140–147.
Wang L., Horiuchi T., Sekimoto A., Okano Y., Ujihara T., Dost S., Three-dimensional numerical analysis of Marangoni convection occurring during the growth process of SiC by the RF-TSSG method. Journal of Crystal, Vol.520 (2019), pp.72–81.
Yamamoto T., Okano Y., Ujihara T., Dost S., Global simulation of the induction heating TSSG process of SiC for the effects of marangoni convection, free surface deformation and seed rotation, Journal of Crystal Growth, Vol.470 (2017), pp.75–78.