


Bayesian Optimization of Terahertz Quantum Cascade Lasers

Martin Franckié^{*} and Jérôme Faist

Institute for Quantum Electronics, ETH Zurich, Auguste-Piccard-Hof 1, 8093 Zurich, Switzerland

 (Received 2 December 2019; revised manuscript received 10 February 2020; accepted 18 February 2020; published 10 March 2020)

We use Bayesian optimization algorithms in combination with a nonequilibrium Green's function transport model to increase the maximum operating temperature of terahertz quantum cascade lasers. This procedure lead to the recent temperature record of 210 K in terahertz quantum cascade lasers, and here we provide even-further-improved structures. The Bayesian optimization algorithm, which takes into account all the available history of the optimization, converges much faster and more securely than the commonly used genetic algorithm. Designs based on two and three wells per period are considered, and using the large amount of data generated, we systematically evaluate the studied schemes in terms of optimal extraction energy and relevance of electron-electron correlations. This analysis shows that the two-well scheme is superior for reaching high operating temperatures, while the three-well scheme is more robust to variations in layer thicknesses. Furthermore, we study the sensitivity to flux-rate fluctuations during growth and simulation-model inaccuracies, showing the period thickness needs to be controlled to within a few percent, which is challenging but achievable with present-day molecular-beam epitaxy. These limits to the growth accuracy can be a guiding principle for experimentalists, along with the suggestion to fabricate devices across the wafer radius so as to find the optimal period thickness.

DOI: [10.1103/PhysRevApplied.13.034025](https://doi.org/10.1103/PhysRevApplied.13.034025)

I. INTRODUCTION

Many quantum phenomena occur at the energy scale of the terahertz photon. Examples include elementary chemical processes [1], photosynthesis [2], superconducting gaps in type-II superconductors [3], (intra)molecular vibrations and phonons in solids, thermal energies, and plasma frequencies in doped semiconductors and metals. In addition, there are several advantages for imaging and spectroscopic applications in the terahertz region [4]. Unlike the nearby mid-infrared spectral region, where quantum cascade lasers (QCLs) [5] have been demonstrated to be compact, powerful, efficient, and relatively cheap sources operating at room temperature, the terahertz region still suffers from a lack of such a source. The two-most-successful terahertz-QCL [6] schemes to date, whose band structures are shown in Fig. 1, are based on two [7] and three [8] quantum wells per period and recently reached an operation temperature of 210 K [7] with thermoelectric cooling [7,9]. While this is a promising step forward, the performance at high temperature still needs to be improved so that high output powers, eventually in continuous-wave mode, can be reached with thermoelectric cooling.

To push terahertz QCLs to their maximum capabilities, automated optimization is needed in combination with accurate simulations.

A wide variety of numerical models for simulating electron transport and light-matter interaction in such heterostructures exist [10]. Models based on rate equations or simplified density-matrix schemes [11,12] can simulate mid-infrared QCLs [13], where the transport is mostly incoherent, with high accuracy and efficiency. This has allowed optimization schemes based on genetic algorithms (GAs) to be applied [14–17] also in the terahertz range [18,19], albeit with less accuracy. Such models are less reliable for simulating terahertz QCLs since coherences play a much bigger role for the transport [20]. Here more advanced models based on nonequilibrium Green's functions (NEGFs) [21–23], ensemble Monte Carlo methods [24,25], or full density matrices [26,27], which treat all elements of the density matrix on the same footing, should be used. Naturally, these more-general models are more computationally demanding, taking up to several hours to evaluate the performance of a single structure, which explains why they have not been used for optimization of QCL structures. For such expensive merit functions, genetic optimization schemes may require too many iterations to converge to the global optimum in a reasonable time. Probabilistic optimization schemes based on Bayesian inference [28] can overcome this limitation by making efficient use of all the available information, and has already proven useful for similar design problems, e.g. concerning thermoelectrics [29,30] and optical metamaterials [31,32].

^{*}martin.franckie@phys.ethz.ch

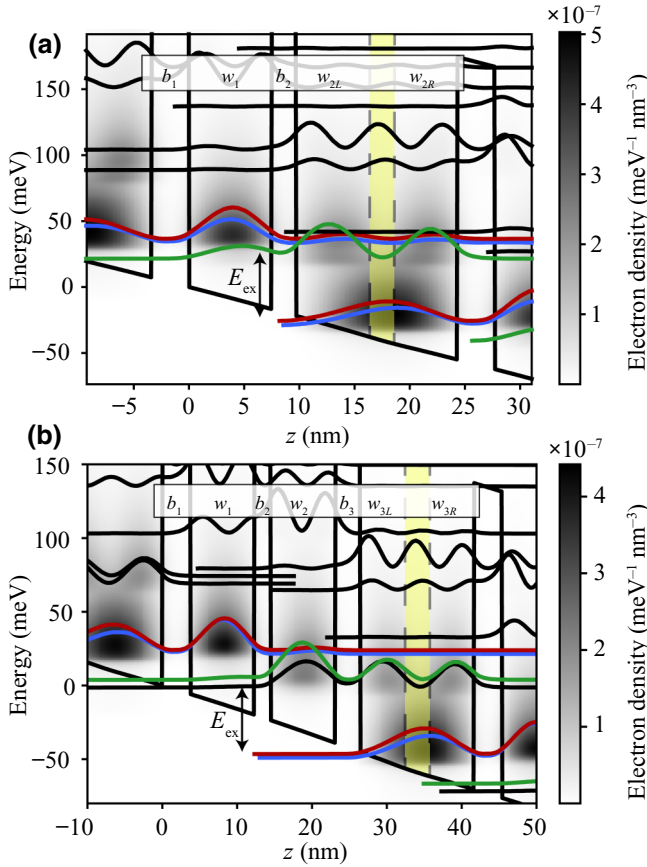


FIG. 1. Band structure (straight black lines), square of the Wannier-Stark functions (wavy lines), and energy-resolved electron densities (gray scale) for (a) the two-well scheme (405 in Table I) with layer sequence 7.507/2.206/6.986/3/4.565/3.425 and (b) the three-well scheme (referred to as 477 in Sec. III) with layer sequence 8.59/3.35/4.81/5/5.40/3.78/8.48/2.22 at $T_L = 300$ K and maximum current density. Labeling of the layers within one period is provided, which together with the barrier composition makes up six (eight) varying parameters for the two-well (three-well) scheme. The extraction energy $E_{\text{ex}} \approx 50$ meV is greatly above the optical phonon energy $E_{\text{LO}} \approx 37$ meV. The upper (red) and lower (green) laser states as well as the injector state (blue) are highlighted, and the yellow-shaded regions indicate doping layers, whose widths are kept fixed throughout the optimization.

In this work, we use a Gaussian-process (GP) optimization algorithm [33] to optimize terahertz QCLs based on design schemes with two and three quantum wells per period using a nonequilibrium Green’s function model [21]. In addition to being a powerful optimization tool, the Bayesian regression technique allows in-depth analysis of the high-dimensional parameter space. By training a GP regression model on all the evaluated QCL structures and creating a “map” of the parameter landscape, we can conduct a general analysis of, and comparison between, the two schemes.

II. OPTIMIZATION SCHEMES

The problem at hand is demanding: minimizing an expensive merit function in high-dimensional parameter space. Therefore, we use techniques that reduce the number of iterations at the cost of some additional computational burden in between iterations. One can broadly divide these into two categories by the way they handle the high dimensionality of the parameter space, namely, one-dimensional and multidimensional schemes. The former category uses linearization of the space onto a space-filling curve (e.g., a Hilbert curve). For the optimization problem reduced to one dimension, there are a multitude of optimization schemes, which we narrow down to methods that benefit from being parallelized and that reduce the number of iterations by taking into account (possibly) the full history of the minimization. For clarity, we consider only one, the “information algorithm with parallel trials” (IAPT) [34], which is explained in Appendix B. This algorithm is based on Bayesian regression, and estimates the probability of the minimum lying within each subinterval. The intervals are ranked according to this probability, and the N_{gen} highest-ranked points in each generation are chosen to be concurrently evaluated in the following iteration. There is a single model parameter r that controls the trade-off between the convergence rate and the global-minimum success rate. The main benefit of this scheme is that the evaluation of the next optimal parameters is fast, and scales only with approximately $\mathcal{O}(N_t)$, where N_t is the number of (historical) training data points, independently of the dimensionality of the problem.

We also consider a multidimensional GP optimization scheme [33,35,36], which is detailed in Appendix C. In this scheme, the merit-function values $\mathbf{y}^*(\mathbf{x}^*)$ are considered as drawn from a normal distribution:

$$p(\mathbf{y}^*|\mathbf{x}^*, \mathbf{x}_t, \mathbf{y}_t, \theta) \sim \mathcal{N}(\mathbf{x}^*|\mu, K), \quad (1)$$

with mean μ and covariance matrix K , where θ is a set of hyperparameters (analogous to r in the IAPT). Here the parameters \mathbf{x} and merit-function values \mathbf{y} from all previous iterations can be considered a labeled training data set. The model is trained on the training data by our optimizing the hyperparameters with respect to the marginal likelihood of the merit-value data given the training parameters using Bayes theorem:

$$p(\theta|\mathbf{y}, \mathbf{x}) \propto p(\mathbf{y}|\mathbf{x}, \theta) \times p(\theta). \quad (2)$$

With use of the mean μ and the variance (the diagonal elements of K), the best points to be evaluated in the next iteration are chosen by a utility function, such as estimated improvement (EI), which we use in this work. This method scales at least as approximately $\mathcal{O}(N_t^3)$ (since it involves inverting the $N_t \times N_t$ covariance matrix K). In our case, training the model on approximately 1700 data points with

eight dimensions takes approximately 120 s. This puts a practical limit of about 5000–20 000 training points, since the training time becomes much longer than the evaluation time per generation and other methods with slower convergence become more beneficial. Thus, there is still potential for this method to be applied to higher-dimensional and rougher parameter landscapes.

These optimization schemes are compared in Fig. 2, with $N_{\text{gen}} = 10$. Here we also compare them to the commonly used GA, with a crossover rate of 0.5, a mutation rate of 0.5, and a mutation size of 50% for each parameter. The algorithm sorts pairs of parents according to their merit values and provides each pair with three offspring each until N_{gen} new structures have been generated. Details are given in Appendix A. GAs are efficient when the merit-function landscape is complicated and a large number of function evaluations are permissible. They can also converge quickly to an arbitrary local minimum by appropriate settings of the mutation and crossover rates. Ideally, these should be compared for a large number of starting conditions on an actual QCL structure. However,

since the evaluation of the actual QCL merit function takes too long for such a comparison, we chose as a test function a Gaussian-process model that has been trained on the two-quantum-well-QCL and three-quantum-well-QCL data presented in Sec. III.

The Gaussian-process optimization scheme converges much faster than the other ones, and the IAPT scheme is even slower than the GA for the two-well QCL. This can be explained by the linearization, which introduces spurious complexity and noise where the actual multidimensional merit function is smooth. In addition, the Hilbert curve used for the linearization does not cover the entire multidimensional parameter space and a single optimum can be seen as two distinct, but similar, optima along the Hilbert curve. However, the variance of the IAPT scheme is lower than that of the GA scheme, which is also seen in Fig. 2(b), meaning it more reliably converges to the *global* optimum.

A similar behavior for the convergence rates of the GP and IAPT schemes can be observed in Fig. 5(a), where single instances of an actual QCL optimization using the IAPT and GP schemes are shown; after an initial exploration period of approximately 30 generations, the GP scheme quickly converges, while the IAPT scheme converges much more slowly. In Supplemental Material [37] we find the same trend for a completely different (analytical) function, showing that this is not specific to the QCL merit functions tested.

In the optimizations presented below, our NEGF model is coupled to the Gaussian-process scheme, as explained in Appendix D. The code which used for generating structures, starting the merit-function evaluations and obtaining the results, and performing the optimization has been published online as part of the open-source code AFTERSHOQ [38]. Further details are given in Appendix E.

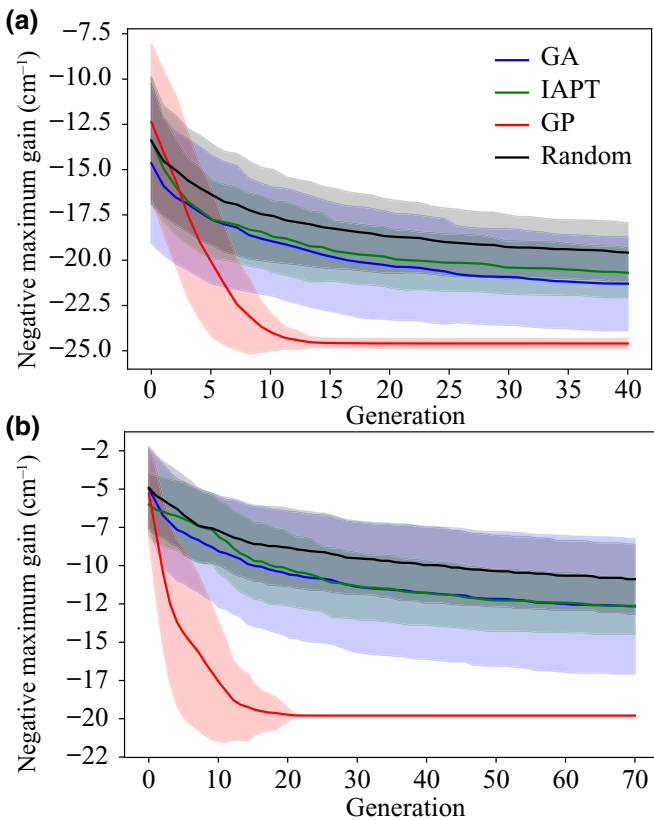


FIG. 2. Comparison of the GA, IAPT, and multivariate GP optimization schemes applied to Gaussian processes trained on (a) two-well-QCL and (b) three-well-QCL simulation data from Sec. III. Each line shows the average of 100 optimization runs, and the shaded regions indicate 1 standard deviation from the mean.

III. RESULTS

A. Optimization of the two-well terahertz QCL

Recently, our group increased the operating temperature of terahertz QCLs using a design optimized using the IAPT described above and varying the four layer widths to maximize the gain at 300 K. The details of this optimization are given in Supplemental Material [37]. The optimized design had a simulated gain of 25 cm^{-1} , which predicts a much-higher operating temperature than observed experimentally [7], considering total optical losses of about 20 cm^{-1} at 300 K for Cu-Cu double-metal waveguides [39]. However, these simulations did not consider the effect of electron-electron (e - e) correlations, which act to redistribute the carriers (thermalization) and greatly reduce the peak gain due to level broadening [40]. When this effect is included [41], the gain is almost halved at a given temperature, reaching only 14 cm^{-1} at room temperature (see Supplemental Material [37]). It is thus clear that e - e correlations are crucial for the device performance

at these doping levels and including them during optimization will allow possible strategies to mitigate their detrimental effects—which are quite different from those of LO phonons—to be explored. (Later we show that this is not the case.) Therefore, we perform optimization for the same two-well structure including e - e correlations and, in addition to all layer widths, the doping position within the wide well and the AAs concentration in the barriers are varied. Figure 1 presents one of the best two-well structures found, and indicates the parameters that are varied during the optimization. As a merit function, the overall maximum gain of the structure is chosen, and is evaluated by our first finding the maximum point on the I - V curve [shown in Fig. 3(b) for the best-five structures] and then the maximum gain as a function of frequency [Fig. 3(a)] at the corresponding bias (unlike the previous optimization, where the gain was evaluated at a predefined bias point).

The maximum gain for the best-five structures obtained with the IAPT optimization scheme are shown in Fig. 3(a).

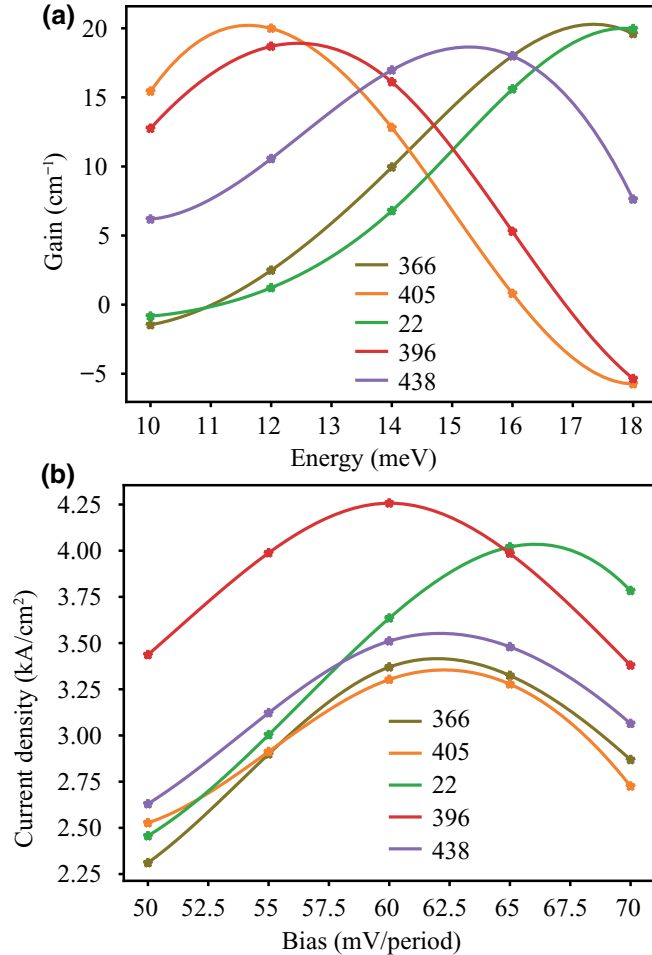


FIG. 3. (a) Gain and (b) current density versus bias for the best-five structures. Markers indicate evaluated points and the solid lines are cubic interpolations used to find the maximum of the I - V and gain curves.

TABLE I. The label, layer sequence with barriers in boldface and doped layers underlined, barrier AAs fraction x , oscillator strength f_{osc} , and maximum current density J_{max} for the best-five two-well structures.

ID	Layer sequence (nm)	x	f_{osc}	J_{max} (A/cm ²)
22	7.286/ 1.897 /7.424/3/4.565/ 3.335	0.268	0.11	4.0
366	7.728/ 1.786 /8.352/3/4.565/ 3.245	0.279	0.17	3.4
396	7.728/ 2.049 /7.424/3/4.698/ 3.065	0.268	0.23	4.2
405	7.507/ 2.206 /6.986/3/4.565/ 3.425	0.276	0.21	3.4
438	7.507/ 2.154 /6.960/3/5.222/ 3.335	0.262	0.19	3.6

Two classes of structures are found, corresponding to different local minima; number 22 and number 366 with gain peaking around 4.5 THz, and number 405 and number 396 with gain peaking around 3 THz. Remarkably, the small change in layer widths between these two classes, given in Table I, results in a drastic change in J_{max} , as seen in Fig. 3(b).

The optimization described increases the gain from 14 to 20 cm⁻¹ at 300 K. The reasons for this increase can be revealed by looking at Fig. 4, where the dependence of the gain on pairs of parameters is shown. The parameter values for the original design are marked by a green asterisk. The main difference from the original design is a much-narrower phonon well (w_{2R} and w_{2L}) as seen in Fig. 4(a), giving a higher extraction energy as high as $\Delta E_{\text{ex}} = 51$ meV. This affects the gain in four ways:

- It limits the phonon emission from the upper laser state.
- It reduces the influence of the reabsorption dip in the gain spectrum at E_{ex} (as defined in Fig. 1), whose width $\hbar/\tau \approx 8$ meV becomes considerable compared with the lasing energy at 300 K. Conventional terahertz QCLs with $E_{\text{ex}} \approx E_{\text{LO}}$ have $E_{\text{ex}} - \hbar\omega \approx 20$ meV—less than $k_B T = 25$ meV at 300 K. In contrast, for the optimized structure $E_{\text{ex}} - \hbar\omega \approx 35$ meV, which is much higher than the thermal energy and any level broadening.
- It reduces the thermal backfilling from the injector state into the lower laser state.
- It increases the lifetime of both the upper laser state and the lower laser state, reducing the broadening and thereby increasing the peak gain.

We also note the best perforation is obtained with the doping position to the right of the well center, where the lower laser state has its node, minimizing the scattering rate into the lower laser state.

For the optimal value for the phonon-well width, the laser-well width (w_l) controls the alignment between the injector state and the upper laser state and thereby the energy drop per period, which is distributed between the photon energy and the extraction energy. This can be seen

by the structure with the narrowest laser well (number 22) having the highest operating bias, while the one with the widest laser well (number 396) has a much-lower operating bias, as seen in Fig. 3(b). As seen in Fig. 4(b), our optimization favors a smaller w_l compared with the nominal design, which is needed to compensate the higher extraction energy.

As seen in Fig. 4(c) the barriers are slightly thinner for the best design, although the laser barrier width b_1 has a small influence on the gain as can be seen also in Figs. 4(d) and 4(e). As shown in Supplemental Material [37], the trained Gaussian-process model predicts that similar or even slightly higher maximum gain is achievable for a range of barrier combinations both thinner and thicker than the nominal design. By changing the parameter values along the ridges of highest gain, we can obtain similar gain values in the whole range from 2.7 to 4.4 THz, as evident in Fig. 3(a). This also allows the oscillator strengths to vary widely from 0.11 and 0.23, as seen by the corresponding values in Table I, which are evaluated at J_{\max} , for the best-five structures. These designs are thus more diagonal than the previous record three-well terahertz QCL [42]. Because

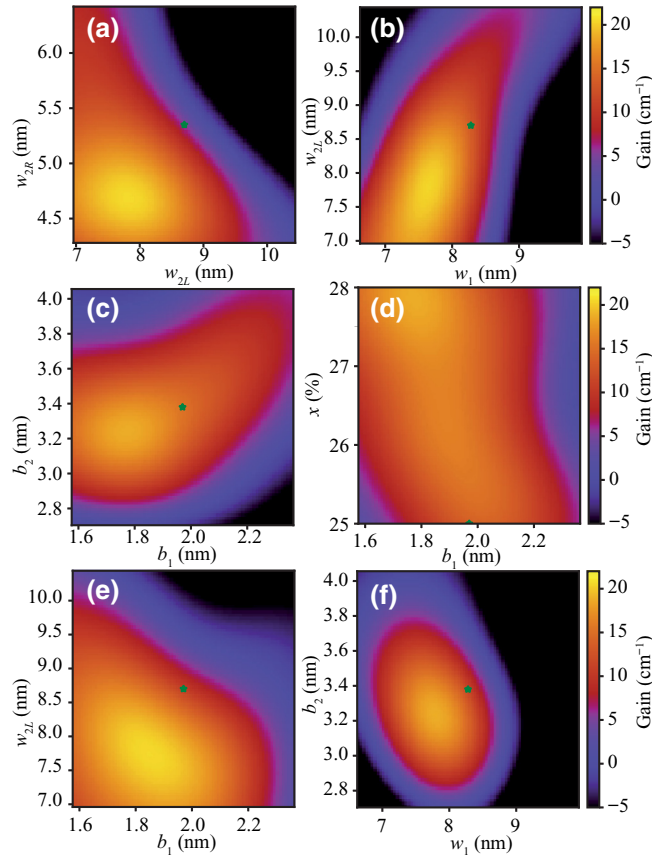


FIG. 4. Gain dependence of the two-well-QCL scheme on pairs of parameters deviating from their optimal values. The green asterisks indicate the nominal-design [7] parameters. The labels are defined in Fig. 1(a).

of the effect of phonon emission from the upper laser state as discussed above, a higher diagonality will likely lead to better performance at very high temperature (300 K) at the cost of worse performance at lower temperatures.

Finally, higher barriers (larger x) improve the performance [see Fig. 4(d)] by reducing carrier leakage into higher-energy states, although the trend is much weaker than for the layer widths. However, the added degree of freedom of one additional variable parameter certainly helps to fine-tune the alignment of the energy levels.

From comparison of the radically different optimal extraction energy at low temperature (approximately E_{LO}) and that at 300 K (approximately 50 meV), it becomes clear that the intended operation temperature is crucial when one is designing structures. This message is important also in other cases, such as low-dissipation or high-power devices operating at lower temperatures. In contrast, we find that $e-e$ scattering, while crucial for the actual performance, is of less importance for designing structures, since the best structures show the relative best performance also without $e-e$ scattering; in contrast to phonon scattering, $e-e$ scattering is not temperature dependent. This indicates that the largest part of the improvement comes from a reduction of the phonon scattering rate out of the upper laser state. Apparently, no strategy for reducing the effect of $e-e$ scattering is found, which means this effect is of lower importance for the relative performance of terahertz QCLs. This is good news since ignoring it allows faster computations with more-easily-interpreted results.

We observe a strong correlation between the widths of the phonon well (w_{2L} and w_{2R}) and well 1, as well as the two barrier widths b_1 and b_2 , as diagonal features in Figs. 4(b) and 4(c). These correlated parameters need to be varied synchronously when one is optimizing the structure. In contrast, the well and barrier widths are much less correlated [see Figs. 4(e) and 4(f)], meaning they control separate features (i.e., energy separation and bias per period for the wells, oscillator strength and tunneling efficiency for the barriers) of the QCL, and can be tuned independently. The high degree of correlation observed in Fig. 4 is a characteristic feature of the two-well scheme, where the layers across the period are connected via the wave functions, which extend over the whole period. This poses a challenge when one is optimizing such short structures since all layers have to be simultaneously and precisely tuned.

B. Optimization of the three-well terahertz QCL

To make a fair comparison between the two-well scheme and the three-well scheme, we optimize the latter with respect to the maximum gain using similar simulation conditions,; that is, varying all eight layer widths and the composition at $T_L = 300$ K. To increase the speed and

improve the convergence of the optimization, we ignore e - e scattering in accordance with the discussion above. The effect of e - e scattering can be checked for the optimized designs afterward (and is found not to change the overall conclusions). Because of the large number of parameters, we use both the IAPT and the multivariate Gaussian-process optimization algorithm to ensure convergence. In addition, a second GP optimization with extended parameter ranges is needed to find the optimal structure. As can be seen in Fig. 5(a), the latter converges much more quickly, confirming the results in Fig. 2(b). In total, approximately 1700 structures are evaluated during these three optimization runs.

In Fig. 5 the gain and current densities of the best-five evaluated designs are shown. A maximum gain of 21 cm^{-1} is achieved for structure 477. Including e - e scattering, we observe a large increase in the current density [circles in Fig. 5(b)] and a large reduction in gain [dashed lines in Fig. 5(c)] to a maximum of 9 cm^{-1} .

In Fig. 6 the maximum gain predicted by the Gaussian-process regression model trained on simulated structures is shown for variation of pairs of parameters away from the optimal values. Compared with the nominal-structure parameters (indicated by a green asterisk in each subplot), we find that the main differences leading to the optimal design are a narrower phonon well (w_{3L} and w_{3R} , as for the two-well scheme) and narrower barriers b_1 and b_3 . Similarly to the two-well scheme, the narrower left portion of the phonon well pushes the extraction energy close to 50 meV. In contrast, the behavior of the doping position is quite the opposite, as the doping is shifted toward the left side of the well, away from the nodes of the lower laser and extractor states. Indeed, since the overlap with the upper laser state of the same period is negligible, a higher impurity scattering rate is beneficial in this case since it reduces the dwell time for electrons in the lower laser state. The narrower barriers provide faster injection and extraction rates out of and into the respective laser levels, and are thus beneficial for inversion, at the cost of a slight gain broadening, which is, however, small in comparison with the lifetime broadening at this high temperature.

As also seen for the two-well scheme, the laser barrier (b_2), which controls mainly the oscillator strength, has only a weak influence on the gain as seen in Figs. 6(c) and 6(d). This suggests that the oscillator strength is not important in this case; the benefit of a higher oscillator strength for the gain apparently cancels to a high degree with the reduced inversion it also brings [43].

IV. DISCUSSION

The Gaussian-process regression model, which is trained on the aggregated QCL data, can be used to draw conclusions about the sensitivity of each design scheme to parameter variations. For clarity, we assume that the

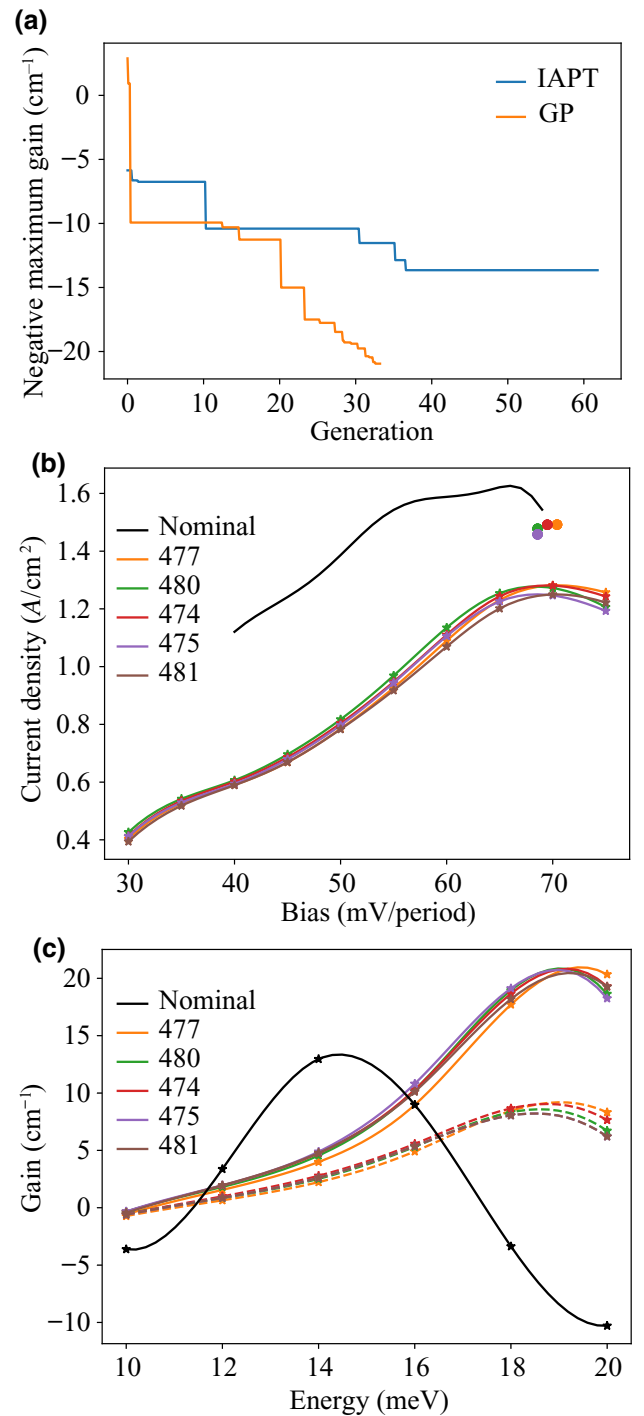


FIG. 5. (a) Convergence rate for the three-well scheme, starting from the structure in Ref. [8]. (b) Current density versus bias and (c) gain for the best-five three-well structures optimized at a lattice temperature of 300 K. In (b),(c), circles and dashed lines, respectively, give the corresponding values with e - e scattering taken into account.

regression represents the true gain as a function of all parameters. The standard deviation in maximum gain ranges from 0 to 3 cm^{-1} for the range of parameters in Figs. 4 and 6.

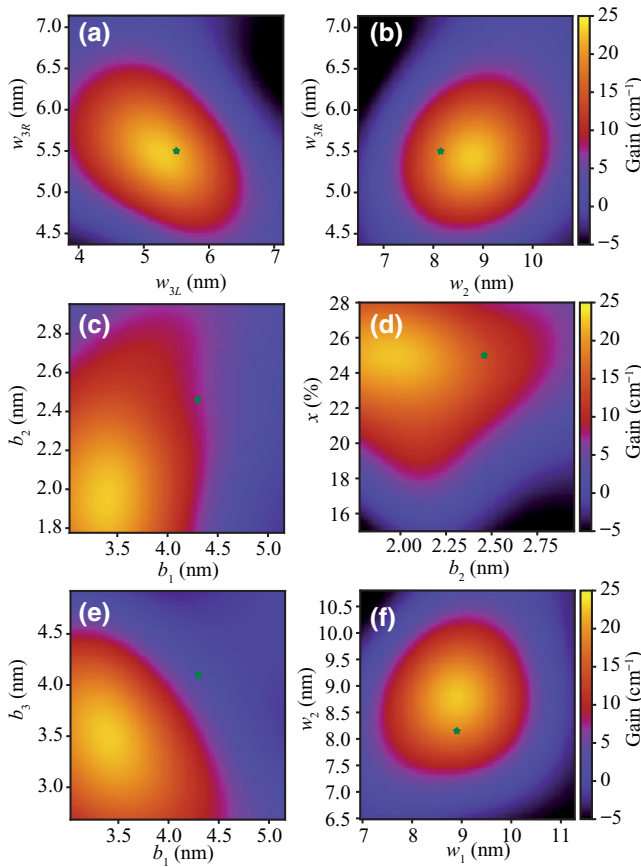


FIG. 6. Gain dependence of the three-well-QCL scheme on pairs of parameters deviating from their optimal values. The green asterisks indicate the nominal-design [8] parameters. The labels are defined in Fig. 1(b).

A first insight can be obtained by comparing Figs. 4 and 6. It is clear that the three-well scheme has many fewer correlations among the layer thicknesses. This indicates that this scheme can be optimized by improving each parameter separately, and that it is less sensitive to variations in individual parameters. Conversely, the two-well design can be improved only by taking into account all wells or barriers simultaneously. This makes it inherently harder to design for a desired feature (such as high-temperature operation).

In comparison with the two-well scheme, the gain including $e-e$ correlations is much lower for the best three-well structure. This strongly supports the argument that the two-well scheme, using only three active levels per period, is superior for high-temperature operation [7]. However, some of the designs have very low current density compared with the two-well designs, as seen in Fig. 5(c). These simulations are performed for a constant sheet doping density, and the gain-current ratio can be tuned to some degree by increasing or decreasing doping. The three-well scheme might therefore be advantageous for

continuous-wave operation, although future optimizations in this regard for the two-well scheme might also yield great improvements, since the ones presented in this work completely ignore the current density.

When one is designing QCLs there are two main concerns: the accuracy of the simulation model and the accuracy of the material fluxes during growth. To address the first one, we may consider inaccuracies in the simulation model as each parameter w_i, x of the predicted optimal design having a random deviation from the actual optimal design parameters. Sampling a large number of structures from the GP and plotting their gain versus the distance from the actual optimal design

$$D^2 = \sum_i (w_i - w_0)^2 \quad (3)$$

(excluding the composition x) is equivalent to sampling along the radius of the N -dimensional hypersphere centered at the respective optimal designs. The sensitivities of each design are shown in Fig. 7(a). In both data sets, the layer widths are varied by 20%, and this definition of the distance excludes any effect from different absolute parameter ranges for the two design schemes. This comparison shows that the two-well scheme has a potentially larger gain (only the two-well simulations include $e-e$ scattering), while the three-well design is more robust to variations in layer widths.

With regard to growth-rate inaccuracies, a fractional change of α in the Ga flux induces width changes in wells (Δw_w) and barriers (Δw_b) of

$$\Delta w_w = \alpha w_w, \quad (4)$$

$$\Delta w_b = \alpha(1 - x)w_b. \quad (5)$$

(The change in composition is negligible. A similar change in the Al flux changes the barrier width by only αx and would also be negligible.) The predicted gain as a function of Ga flux is shown in Fig. 7(b) and again shows more robustness for the three-well scheme, and an asymmetry for the two-well scheme as wider layers are more detrimental than thinner ones (as seen in Fig. 4). This shows that the growth rate must be controlled to within a few percent and only half the gain remains for a change of 5.5% (3%) for the three-well (two-well) design. As molecular-beam-epitaxy growth rates typically change by a similar amount across a wafer [7], the performance of three-well-QCL devices is more reliable. These considerations can explain why three-well QCLs rather than two-well QCLs historically have been better performing in experiments. The difference between the two schemes is less pronounced in Fig. 7(b) than in Fig. 7(a), which is expected because of the higher degree of correlation between layer widths for the two-well design (cf. Figs. 4 and 6).

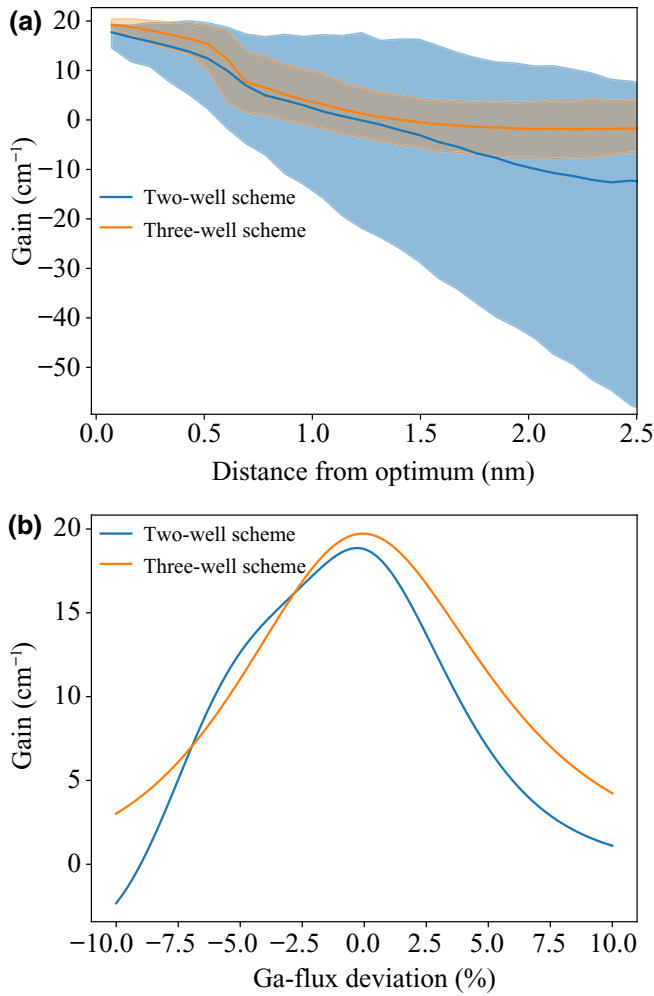


FIG. 7. (a) Posterior mean of the gain averaged over 100 000 binned samples as a function of the radius of a hypersphere of six (eight) dimensions centered at the optimal design for the two-well (three-well) design scheme (lines). The regions between the first percentile and the 99th percentile are shaded. (b) Gain as a function of changes in the Ga flux during molecular-beam-epitaxy growth.

The precision required to reach the highest gain is remarkable. Figure 7(b) indicates the period length needs to be controlled to within 1%. Luckily, this is possible with current molecular-beam-epitaxy technology, where terahertz-QCL structures are now commonly grown with this precision. For instance, Beere *et al.* [44] reproduced the active-region thickness to within 2%, with less than 2% difference in period length along the whole 2-in. wafer. The thickness deviation along the growth direction is much smaller, as any drift in the growth rate is compensated during growth, indicating a sample with the correct thickness can be found on the wafer, even if the thickness at a particular point on the wafer is wrong. This is precisely what we found in Ref. [7]; spatially resolved high-resolution x-ray-diffraction measurements showed a

thickness deviation of 4% across the 3-in. wafer, which allowed the optimal sample to be found close to the wafer edge.

Finally, we mention that e - e scattering is excluded for optimization of the three-well design, which means that the optimal design found might not precisely correspond to the best three-well design if e - e scattering were included. However, the overall best design is the two-well design (since e - e scattering degrades the gain of even the best three-well design to below that of the best two-well design), which is optimized by including our e - e scattering. We therefore present an optimal design for high-temperature terahertz-QCL operation within the parameter ranges studied.

V. CONCLUSIONS

In conclusion, we investigate in detail the gain landscape of two-well and three-well QCLs using Gaussian-process regression, which allows a general assessment of respective benefits and deficits of these design schemes. Specifically, we find that the two-well scheme is superior in terms of high-temperature gain, while the three-well scheme shows more robustness toward growth and model inaccuracies. Both design schemes are optimized with use of Bayesian methods, and we find a gain of more than 20 cm⁻¹ at 300 K, even when e - e correlations are included. This indicates pulsed terahertz-QCL operation close to room temperature in a wide range of terahertz frequencies might be possible soon. The gain is increased mainly by a large increase in the extraction energy, far above the LO phonon energy, suggesting a new design feature that can be applied to terahertz QCLs in general to reach higher operating temperatures. Finally, the gain could be further increased by varying also the doping density and the composition of each layer separately, although the former would probably require more experimental feedback as e - e scattering becomes more efficient at higher doping levels.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support from the Qombs Project funded by the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 820419. The work was carried out within the Collaborative Research Centre 956, sub-project D1, funded by the Deutsche Forschungsgemeinschaft (DFG). M.F. thanks Matthias Beck, Giacomo Scalari, and Mathieu Bertrand for fruitful discussions.

APPENDIX A: GENETIC ALGORITHM

The genetic algorithm shown in Fig. 2 is implemented as follows. At the beginning a number of random points in the multidimensional parameter space are selected and

the model is evaluated. The N_{pop} structures with the best merit function are selected for mutation and crossover. In the crossover phase, two parents are chosen at random with probability

$$P_i = \frac{y_i}{\sum_j y_j}, \quad (\text{A1})$$

where i and j are indices going over the N_{parent} best structures. For each pair of parents, N_{off} offspring are created, where each gene (i.e., each parameter) in the offspring is given by $p_{\text{off}} = p_1 w + p_2(1 - w)$, where p_i is the parameter value of parent i and $w \in (0, 1)$ is a random weight.

In the mutation phase, every gene of each offspring has a probability r_{mut} changed by a random amount $\Delta p_{\text{mut}} = \kappa(p_{\text{max}} - p_{\text{min}})$, where $\kappa \in (-m, m)$ is a random number in the range of the mutation size parameter $m \in (0, 1)$. The remaining $N_{\text{pop}} - N_{\text{off}}N_{\text{parent}}$ structures are brought over to the next generation unchanged.

APPENDIX B: INFORMATION ALGORITHM WITH PARALLEL TRIALS

This algorithm [34] finds the global minimum of a one-dimensional function, accelerated by using parallel function evaluations. The intervals between each pair of evaluated points are ranked according to a characteristic function

$$\begin{aligned} R_i &= x_i - x_{i-1} + \frac{(y_i - y_{i-1})}{\mu^2(x_i - x_{i-1})} - 2\frac{(y_i + y_{i-1})}{r\mu}, \\ R_1 &= 2(x_1 - x_0) - 4\frac{y_1}{r\mu}, \quad R_N = 2(x_N - x_{N-1}) - 4\frac{y_{N-1}}{r\mu}, \\ \mu &= \max \frac{|y_i - y_{i-1}|}{(x_i - x_{i-1})} \quad (1 < i \leq N - 1), \end{aligned} \quad (\text{B1})$$

used to determine the interval that most likely contains the minimum of the merit function. For each of the N highest-ranked intervals, the new x value is given by

$$\begin{aligned} x'_i &= (x_i + x_{i-1})/2 - (y_i - y_{i-1})/2r\mu, \quad 1 < i < N, \\ x'_i &= (x_i + x_{i-1})/2, \quad i = 1 \text{ or } i = N, \end{aligned} \quad (\text{B2})$$

and these values are taken as the next N points to be evaluated. The parameter r determines the amount of exploration and exploitation. The characteristic functions R_i are the maximum-likelihood estimates of each interval containing the global minimum, and their derivation is based on Bayesian reasoning [45].

Since the algorithm is applicable only for one-dimensional functions, the multidimensional parameter space has to be transformed into a one-dimensional one. This is done by the use of a Hilbert curve, a one-to-one mapping from M to one dimension, which preserves relative distances, (i.e., two points that are close on the Hilbert

curve are also close in real space, but not necessarily vice versa). Thus, not all points in real space are represented on the Hilbert curve. By the choice of the only Hilbert-curve parameter p , determining the minimum distance between two parallel sections of the curve, every point in real space can be made sufficiently close to a point on the Hilbert curve.

APPENDIX C: GAUSSIAN-PROCESS REGRESSION

Gaussian-process regression is a powerful tool for analyzing large sets of data, and can be used for optimization. In essence, the process attempts to fit a statistical distribution with a set of hyperparameters to the input data, providing both the most-likely (mean) value and also information on the statistics via the full covariance matrix. This is done by defining a Gaussian-type kernel k of the covariance matrix:

$$k(x, x') = \sigma_f^2 \exp(-\frac{1}{2}\sigma_l^2|x - x'|^2). \quad (\text{C1})$$

The parameters σ_f and σ_l are the hyperparameters, which can be fitted by maximizing the posterior probability of σ_f and σ_l given the input data, as shown below. Consider two sets of data points: the already evaluated N_p prior points \mathbf{x} (the training data) and the N_t test points \mathbf{x}^* (where $N_t \gg N_p$). The covariance matrix thus becomes

$$\Sigma = \begin{pmatrix} \mathbf{K} & \mathbf{k}^* \\ \mathbf{k}^{*T} & \mathbf{k}^{**} \end{pmatrix}, \quad (\text{C2})$$

where \mathbf{K} has the kernel between points in the training data, \mathbf{k}^* has the kernel between the training and the test data, and \mathbf{k}^{**} has the kernel between points in the test data. It is well known (see, e.g., Ref. [46]) that if the prior likelihood distribution of the data given the parameters

$$p(\mathbf{y}|\mathbf{x}, \sigma_f, \sigma_l) = \mathcal{N}(\mathbf{x}|E(\mathbf{x}), \Sigma) \quad (\text{C3})$$

is a Gaussian distribution, then the posterior distribution of the new data conditional on the training data

$$\begin{aligned} p(\mathbf{y}^*|\mathbf{x}^*, \mathbf{x}, \mathbf{y}, \sigma_f, \sigma_l) &= \frac{p(\mathbf{y}^*, \mathbf{x}^*, \mathbf{x}, \mathbf{y}, \sigma_f, \sigma_l)}{p(\mathbf{y}|\mathbf{x}, \sigma_f, \sigma_l)} \\ &= \mathcal{N}(\mathbf{x}^*|\mu', \Sigma') \end{aligned} \quad (\text{C4})$$

is also a Gaussian distribution with mean $\mu' = \mathbf{k}^{*T}\mathbf{K}^{-1}\mathbf{y}$ and covariance $\Sigma' = \mathbf{k}^{**} - \mathbf{k}^{*T}\mathbf{K}^{-1}\mathbf{k}^*$. The mean will thus be a smooth function passing through all training data points, and can be seen as a fit with uncertainty given by the variance $\Sigma'(\mathbf{x}^*, \mathbf{x}^*)$. This fit strongly depends on the choices of σ_f and σ_l , which are optimized with respect to the training data by maximizing the posterior probability

of σ_f and σ_l given the training data, which by the Bayes theorem is

$$p(\sigma_f, \sigma_l | \mathbf{y}, \mathbf{x}) = p(\mathbf{y} | \mathbf{x}, \sigma_f, \sigma_l) \times p(\sigma_l, \sigma_f). \quad (\text{C5})$$

It is enough to maximize $p(\mathbf{y} | \mathbf{x}, \sigma_f, \sigma_l)$, and using Eq. (C4), we find

$$\log p(\mathbf{y} | \mathbf{x}, \sigma_f, \sigma_l) = -\frac{1}{2} \mathbf{y}^T \mathbf{K}^{-1} \mathbf{y} - \frac{1}{2} \log |\mathbf{K}| - \frac{n}{2} \log 2\pi. \quad (\text{C6})$$

Together with its derivative

$$\frac{\partial}{\partial \theta} \log p(\mathbf{y} | \mathbf{x}, \theta) = \frac{1}{2} \mathbf{y}^T \mathbf{K}^{-1} \frac{\partial \mathbf{K}}{\partial \theta} \mathbf{K}^{-1} \mathbf{y} - \frac{1}{2} \text{Tr} \left(\mathbf{K}^{-1} \frac{\partial \mathbf{K}}{\partial \theta} \right) \quad (\text{C7})$$

[where $\theta = (\sigma_f, \sigma_l)$] we can use a built-in serial solver to maximize the probability function, since this is a much-easier problem than evaluating the merit function.

In the above derivation, we assume that the evaluated data have no uncertainty (since the NEGF program will always yield exactly the same result this is a logical choice). If the model result is associated with some uncertainty (such as Monte Carlo models, or considering some uncertainty in the model parameters, such as interface roughness), noise (with variance σ_N^2) can be included in the evaluated data by addition of a term $\hat{I}\sigma_N^2$, where \hat{I} is the $N_p \times N_p$ identity matrix, to the kernel \mathbf{K} of the training data. This can also increase the convergence rate, since fast oscillations (as those observed on the Hilbert curve) are averaged out as the Gaussian-process mean no longer has to go through each training point.

The first step in the optimization is to choose initial points and evaluate the model at those points (typically ten to 20 points, either randomly chosen to be on the Hilbert curve or directly in multidimensional parameter space). Initial values of σ_f , σ_l , and σ_N are provided and the Gaussian process is trained on the data. A utility function determines the next M points to be evaluated, using the information on both the mean and the variance at each parameter-space point. Here a wide variety of choices are available and can greatly affect the convergence rate and robustness of the optimization. For example, the utility function

$$u_{\text{ML}}(\mathbf{x}^*) = -\mu'(\mathbf{x}^*) + \sigma(\mathbf{x}^*), \quad (\text{C8})$$

where $\sigma(\mathbf{x}^*) = \sqrt{\text{diag} \Sigma'}$, is the maximum likelihood with an additional reward for points with high uncertainty, preventing the convergence to local minima. To select multiple points at different regions of parameter space, a minimum distance between chosen points is also specified

in this case. Another choice is the estimated-improvement utility function:

$$u_{\text{EI}}(\mathbf{x}^*) = \mathbb{E}[\mathbf{y}^*(\mathbf{x}^*) - \mathbf{y}_{\text{max}}] = [\mu'(\mathbf{x}^*) - \mathbf{y}_{\text{max}} - \xi]F(Z) + \sigma(\mathbf{x}^*)P(Z), \quad (\text{C9})$$

where F is the cumulative distribution function, P is the probability density function, $\mathbf{y}_{\text{max}} = \max \mathbf{y}$ and

$$Z = \frac{\mu'(\mathbf{x}^*) - \mathbf{y}_{\text{max}} - \xi}{\sigma(\mathbf{x}^*)}. \quad (\text{C10})$$

The parameter $\xi = 0.01$ determines the amount of exploration versus exploitation. For more details on the precise implementation in the multidimensional case, see the open-source code in Ref. [33].

Once the next parameter points have been selected, the model is evaluated for them and the Gaussian process is retrained on the updated data set. This procedure is repeated until convergence, or until the maximum number of generations has been reached. We use this algorithm in both the multidimensional case and the one-dimensional case (i.e., using a Hilbert curve to reduce dimensionality), and a comparison is provided in Supplemental Material [37].

APPENDIX D: NONEQUILIBRIUM GREEN'S FUNCTION MODEL

As supported by the results in Sec. IV, the accuracy of the model is crucial given the small tolerance of layer thicknesses. We use the nonequilibrium Green's function presented in Ref. [21], extended with a two-band model according to Ref. [47], which has proven accurate with respect to experiments for both mid-infrared QCLs [13] and terahertz QCLs [40,48]. As input to the model, the precise layer sequence and barrier composition are provided. The composition x gives the conduction-band offset ΔE_c between wells and barriers according to the equation

$$\Delta E_c = 0.831x \text{ eV} \quad (\text{D1})$$

as suggested in Ref. [49]. The width parameters give the average interface positions z_i . The precise values $z_i(\mathbf{r})$ vary with the in-plane coordinate \mathbf{r} , which is included in the model via interface roughness scattering. In the simulations presented, this is implemented with an exponential autocorrelation function of the deviation $\eta(\mathbf{r})$ from z_i :

$$\langle \eta(\mathbf{r}) \eta(0) \rangle = \Delta^2 e^{-|\mathbf{r}|/\lambda}, \quad (\text{D2})$$

where $\Delta = 0.1$ nm is the rms roughness height and $\lambda = 10$ nm is the correlation length. For each set of parameters, the Wannier functions of the periodic heterostructure are calculated, and the current density at a few bias points is

calculated to find the peak of the current-bias curve, where then the gain is evaluated. To reduce the total computation time, a few points are evaluated with the NEGF model and are then interpolated with a spline function. The highest point of the interpolated gain curve is then chosen as the merit function and is fed into the optimization algorithm.

Since the two-well structure is short, we find that two neighboring periods (on each side of the central period) are needed to achieve an accurate gain calculation, whereas for the three-well structure, one neighboring period is sufficient. In both cases, seven states per period are included in the calculations.

APPENDIX E: SAMPLE CODES

Sample codes (PYTHON JUPYTER notebooks) are available from Ref. [38] under `examples/publi/2020/`, including one script for loading the trained Gaussian-process models used in Secs. II and III (`opt_schemes_trained_GP.ipynb`) and one script used to run the actual optimizations of the two-well QCL using the Gaussian-process regression (`opt_2-well_sequential_MDGP_e-e.ipynb`). As of now, the NEGF model has not been made available to the public. However, the open-source code `AFTERSHOQ` can be extended by the interested reader to include interfaces with other transport simulation codes as well by writing an appropriate interface code accepting the specified inputs and outputting the simulation results in the correct format.

-
- [1] T. Kiwa, T. Kamiya, T. Morimoto, K. Fujiwara, Y. Maeno, Y. Akiwa, M. Iida, T. Kuroda, K. Sakai, H. Nose, M. Kobayashi, and K. Tsukada, Imaging of chemical reactions using a terahertz chemical microscope, *Photonics* **6**, 10 (2019).
- [2] P. N. Nguyen, H. Watanabe, Y. Tamaki, O. Ishitani, and S.-I. Kimura, Relaxation dynamics of $[\text{Re}(\text{CO})_2(\text{bpy})\{\text{P}(\text{OEt})_3\}_2](\text{PF}_6)$ in TEOA solvent measured by timeresolved attenuated total reflection terahertz spectroscopy, *Sci. Rep.* **9**, 1 (2019).
- [3] G. Blatter, M. V. Feigel'man, V. B. Geshkenbein, A. I. Larkin, and V. M. Vinokur, Vortices in high-temperature superconductors, *Rev. Mod. Phys.* **66**, 1125 (1994).
- [4] P. Jepsen, D. Cooke, and M. Koch, Terahertz spectroscopy and imaging – Modern techniques and applications, *Laser Photon. Rev.* **5**, 124 (2011).
- [5] J. Faist, F. Capasso, D. L. Sivco, C. Sirtori, A. L. Hutchinson, and A. Y. Cho, Quantum cascade laser, *Science* **264**, 553 (1994).
- [6] R. Köhler, A. Tredicucci, F. Beltram, H. E. Beere, E. H. Linfield, A. G. Davies, D. A. Ritchie, R. C. Iotti, and F. Rossi, Terahertz semiconductor-heterostructure laser, *Nature* **417**, 156 (2002).
- [7] L. Bosco, M. Franckić, G. Scalari, M. Beck, A. Wacker, and J. Faist, Thermoelectrically cooled THz quantum cascade laser operating up to 210 K, *Appl. Phys. Lett.* **115**, 010601 (2019).
- [8] S. Fatholouloumi, E. Dupont, C. W. I. Chan, Z. R. Wasilewski, S. R. Laframboise, D. Ban, A. Mátyás, C. Jirauschek, Q. Hu, and H. C. Liu, Terahertz quantum cascade lasers operating up to ~ 200 K with optimized oscillator strength and improved injection tunneling, *Opt. Express* **20**, 3866 (2012).
- [9] M. A. Kainz, M. P. Semtsiv, G. Tsianos, S. Kurlov, W. T. Masselink, S. Schönhuber, H. Detz, W. Schrenk, K. Unterrainer, G. Strasser, and A. M. Andrews, Thermoelectric-cooled terahertz quantum cascade lasers, *Opt. Express* **27**, 20688 (2019).
- [10] C. Jirauschek and T. Kubis, Modeling techniques for quantum cascade lasers, *Appl. Phys. Rev.* **1**, 011307 (2014).
- [11] R. Terazzi and J. Faist, A density matrix model of transport and radiation in quantum cascade lasers, *New J. Phys.* **12**, 033045 (2010).
- [12] E. Dupont, S. Fatholouloumi, and H. C. Liu, Simplified density-matrix model applied to three-well terahertz quantum cascade lasers, *Phys. Rev. B* **81**, 205311 (2010).
- [13] M. Lindskog, J. M. Wolf, V. Trinite, V. Liverini, J. Faist, G. Maisons, M. Carras, R. Aidam, R. Ostendorf, and A. Wacker, Comparative analysis of quantum cascade laser modeling based on density matrices and non-equilibrium Green's functions, *Appl. Phys. Lett.* **105**, 103106 (2014).
- [14] A. Bismuto, R. Terazzi, B. Hinkov, M. Beck, and J. Faist, Fully automatized quantum cascade laser design by genetic optimization, *Appl. Phys. Lett.* **101**, 021103 (2012).
- [15] A. Daničić, J. Radovanović, V. Milanović, D. Indjin, and Z. Ikonić, Optimization and magnetic-field tunability of quantum cascade laser for applications in trace gas detection and monitoring, *J. Phys. D: Appl. Phys.* **43**, 045101 (2010).
- [16] A. Gajic, J. Radovanovic, V. Milanovic, D. Indjin, and Z. Ikonic, Optimizing optical nonlinearities in GaInAs/AlInAs quantum cascade lasers, *Nucl. Technol. Radiat. Prot.* **29**, 10 (2014).
- [17] J. M. Wolf, Type doctoral thesis, School ETH Zurich (2017).
- [18] M. T. Arafin, N. Islam, S. Roy, and S. Islam, Performance optimization for terahertz quantum cascade laser at higher temperature using genetic algorithm, *Opt. Quantum Electron.* **44**, 701 (2012).
- [19] A. Matyas, R. Chashmahcharagh, I. Kovacs, P. Lugli, K. Vijayraghavan, M. A. Belkin, and C. Jirauschek, Improved terahertz quantum cascade laser with variable height barriers, *J. Appl. Phys.* **111**, 103106 (2012).
- [20] H. Callebaut and Q. Hu, Importance of coherence for electron transport in terahertz quantum cascade lasers, *J. Appl. Phys.* **98**, 104505 (2005).
- [21] A. Wacker, M. Lindskog, and D. Winge, Nonequilibrium green's function model for simulation of quantum cascade laser devices under operating conditions, *IEEE J. Sel. Topics Quantum Electron.* **19**, 1200611 (2013).
- [22] T. Kubis, P. Vogl, and Self-consistent quantum transport theory, Applications and assessment of approximate models, *J. Comput. Electron.* **6**, 183 (2007).
- [23] G. Haldaś, A. Kolek, and I. Tralle, Modeling of mid-infrared quantum cascade laser by means of nonequilibrium

- green's functions, *IEEE J. Quantum Electron.* **47**, 878 (2011).
- [24] C. Jirauschek, G. Scarpa, P. Lugli, M. S. Vitiello, and G. Scamarcio, Comparative analysis of resonant phonon THz quantum cascade lasers, *J. Appl. Phys.* **101**, 086109 (2007).
- [25] P. Harrison and R. W. Kelsall, The relative importance of electron–electron and electron–phonon scattering in terahertz quantum cascade lasers, *Solid State Electron.* **42**, 1449 (1998).
- [26] A. Gordon, D. Majer, and Coherent transport in semiconductor heterostructures, A phenomenological approach, *Phys. Rev. B* **80**, 195317 (2009).
- [27] G. Kiršanskas, M. Franckié, and A. Wacker, Phenomenological position and energy resolving Lindblad approach to quantum kinetics, *Phys. Rev. B.* **97**, 035432 (2018).
- [28] U. von Toussaint, Bayesian inference in physics, *Rev. Mod. Phys.* **83**, 943 (2011).
- [29] S. Ju, T. Shiga, L. Feng, Z. Hou, K. Tsuda, and J. Shiomi, Designing Nanostructures for Phonon Transport via Bayesian Optimization, *Phys. Rev. X* **7**, 021024 (2017).
- [30] M. Yamawaki, M. Ohnishi, S. Ju, and J. Shiomi, Multifunctional structural design of graphene thermoelectrics by Bayesian optimization, *Sci. Adv.* **4**, 4192 (eaar2018).
- [31] P. Gutsche, P.-I. Schneider, S. Burger, and M. Nieto-Vesperinas, Chiral scatterers designed by Bayesian optimization, *J. Phys.: Conf. Ser.* **963**, 012004 (2018).
- [32] A. Sakurai, K. Yada, T. Simomura, S. Ju, M. Kashiwagi, H. Okada, T. Nagao, K. Tsuda, and J. Shiomi, Ultranarrowband wavelength-selective thermal emission with aperiodic multilayered metamaterials designed by Bayesian optimization, *ACS Cent. Sci.* **5**, 319 (2019).
- [33] The GPyOpt authors, GPyOpt: A Bayesian Optimization framework in Python (2016).
- [34] Y. D. Sergeev and R. G. Strongin, A global minimization algorithm with parallel iterations, *USSR Comput. Math. Math. Phys.* **29**, 7 (1989).
- [35] C. K. I. Williams, in *Learning in Graphical Models*, edited by M. I. Jordan (Springer, Netherlands, Dordrecht, 1998), p. 599.
- [36] C. E. Rasmussen, in *Advanced Lectures on Machine Learning*, edited by O. Bousquet, U. von Luxburg, and G. Rätsch (Springer Berlin, Heidelberg, 2004), p. 63.
- [37] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevApplied.13.034025> for more details on comparisons of optimization schemes and two-well optimization without e - e scattering.
- [38] M. Franckié, Aftershoq: A Flexible Tool for EM-Radiation-emitting Semiconductor Heterostructure Optimization using Quantum models (2018), <https://github.com/mfranckie/aftershoq>.
- [39] Y. J. Han, L. H. Li, J. Zhu, A. Valavanis, J. R. Freeman, L. Chen, M. Rosamond, P. Dean, A. G. Davies, and E. H. Linfield, Silver-based surface plasmon waveguide for terahertz quantum cascade lasers, *Opt. Express* **26**, 3814 (2018).
- [40] M. Franckié, L. Bosco, M. Beck, C. Bonzon, E. Mavrona, G. Scalari, A. Wacker, and J. Faist, Two-well quantum cascade laser optimization by non-equilibrium Green's function modelling, *Appl. Phys. Lett.* **112**, 021104 (2018).
- [41] D. O. Winge, M. Franckié, C. Verdozzi, A. Wacker, and M. F. Pereira, Simple electron-electron scattering in non-equilibrium Green's function simulations, *J. Phys.: Conf. Ser.* **696**, 012013 (2016).
- [42] S. Fatholouloumi, E. Dupont, Z. R. Wasilewski, C. W. I. Chan, S. G. Razavipour, S. R. Laframboise, S. Huang, Q. Hu, D. Ban, and H. C. Liu, Effect of oscillator strength and intermediate resonance on the performance of resonant phonon-based terahertz quantum cascade lasers, *J. Appl. Phys.* **113**, 113109 (2013).
- [43] C. W. I. Chan, A. Albo, Q. Hu, and J. L. Reno, Trade-offs between oscillator strength and lifetime in terahertz quantum cascade lasers, *Appl. Phys. Lett.* **109**, 201104 (2016).
- [44] H. E. Beere, J. C. Fowler, J. Alton, E. H. Linfield, D. A. Ritchie, R. Köhler, A. Tredicucci, G. Scalari, L. Ajili, J. Faist, and S. Barbieri, MBE growth of terahertz quantum cascade lasers, *J. Crystal Growth 13th Int. Conf. Mol. Beam Epitaxy* **278**, 756 (2005).
- [45] R. G. Strongin, The information approach to multiextremal optimization problems, *Stochastics Stochastic Rep.* **27**, 65 (1989).
- [46] S. Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective* (Elsevier, AP – Academic Press, USA, 2015).
- [47] M. Lindskog, D. O. Winge, and A. Wacker, in *Proc. SPIE* Vol. 8846 (2013), p. 884603.
- [48] D. O. Winge, M. Franckié, and A. Wacker, Simulating terahertz quantum cascade lasers: Trends from samples from different labs, *J. Appl. Phys.* **120**, 114302 (2016).
- [49] W. Yi, V. Narayanamurti, H. Lu, M. A. Scarpulla, and A. C. Gossard, Probing semiconductor band structures and heterojunction interface properties with ballistic carrier emission: GaAs/Al_xGa_{1-x}As as a model system, *Phys. Rev. B.* **81**, 235325 (2010).