PAPER • OPEN ACCESS

Realistic mask generation for matter-wave lithography via machine learning

To cite this article: Johannes Fiedler *et al* 2023 *Mach. Learn.: Sci. Technol.* **4** 025028

View the article online for updates and enhancements.

You may also like

- Radiation tolerance studies using fault injection on the Readout Control FPGA design of the ALICE TPC detector
  J Alme, D Fehlker, C Lippmann et al.

- Mollow splitting of the Na D fluorescence lines in an Ar-filled vapour cell
  A R D van Bergen, H J van Halewijn, T Hollander et al.

- Hemispheric vs. regional North Atlantic climate change during the past 1200 years: Responses to natural and man made forcings in data and models
  Eystein Jansen, O H Otteraa, R Telford et al.

MACHINE LEARNING
Science and Technology

PAPER

Realistic mask generation for matter-wave lithography via machine learning

Johannes Fiedler[1,*] ⓘ, Adriá Salvador Palau[1], Eivind Kristen Osestad[1] ⓘ, Pekka Parviainen[2] and Bodil Holst[1] ⓘ

1   Department of Physics and Technology, University of Bergen, Allégaten 55, 5007 Bergen, Norway
2   Department of Informatics, University of Bergen, HIB—Thormøhlens gate 55, 5006 Bergen, Norway
*   Author to whom any correspondence should be addressed.

E-mail: johannes.fiedler@uib.no

Keywords: matter-wave lithography, inverse problems, dispersion forces

Abstract

Fast production of large-area patterns is crucial for the established semiconductor industry and enables industrial-scale production of next-generation quantum devices. Metastable atom lithography with binary holography masks has been suggested as a higher resolution/low-cost alternative to the current state of the art: extreme ultraviolet lithography. However, it was recently shown that the interaction of the metastable atoms with the mask material (SiN) leads to a strong perturbation of the wavefront, not included in the existing mask generation theory, which is based on classical scalar waves. This means that the inverse problem (creating a mask based on the desired pattern) cannot be solved analytically, even in 1D. Here we present a machine-learning approach to mask generation targeted for metastable atoms. Our algorithm uses a combination of genetic optimisation and deep learning to obtain the mask. A novel deep neural architecture is trained to produce an initial approximation of the mask. This approximation is then used to generate the initial population of the genetic optimisation algorithm that can converge to arbitrary precision. We demonstrate the generation of arbitrary 1D patterns for system dimensions within the Fraunhofer approximation limit.

1. Introduction

All semiconductor device fabrication is currently based on pattern generation using mask-based photolithography. The aim is to create patterns with smaller and smaller features at higher and higher information densities. The state-of-the-art extreme ultraviolet (EUV) photolithography, which, with photons (electromagnetic waves) of a wavelength of 13.5 nm, should be able to produce patterns with a resolution of 6.75 nm according to the Abbe criterion. Smaller features could potentially be achieved using immersion and/or over-exposure or under-development, as is done with the 193 nm light source, EUVs predecessor. However, the problem is that in EUV lithography, due to the high energy of the photons (91.8 eV), the pattern generation process is mediated by photo-generated secondary electrons, which can travel for several nm before inducing a reaction. Current experiments and theory indicate that the secondary electron blur radius for EUV is around 3 nm, limiting the feature size that can be achieved to around 6 nm [1]. This means devices based on small quantum dots and individual atoms and molecules cannot be produced with EUV. Moving to wavelengths shorter than 13.5 nm would exacerbate the secondary electron issue—an alternative to photons is needed.

Lithography with atoms (matter waves) was proposed as an alternative to photolithography almost three decades ago. As stated in the first paper on this topic from 1995, the motivation remains the same. '*Unlike techniques that use beams of electrons or ions, there are no electrostatic interactions within a neutral beam that limit focusing or flux density [...] Because of their short (< 0.01 nm) de-Broglie wavelength, thermal beams of neutral atoms can in principle be focused (4) to a spot that is limited by the size of the atom.*' [2].

A further advantage of atom lithography is that, for a given wavelength, the kinetic energy of an atom is much less than the energy of a photon. A helium atom with the same wavelength as a EUV photon (13.5 m)

has a kinetic energy of only $E = h^2/(2m\lambda^2) \approx 0.011\,\text{eV}$, where h denotes Planck's constant and m is the mass of the helium atom. This is almost a factor of 10.000 less than the EUV photon. It is too little to modify a resist, which is why metastable atoms are often used; see below. When a metastable atom hits the substrate, it decays, and the decay energy is transferred to the substrate [2–4]. The energy released when a metastable atom decays is about 10 eV for argon and 20 eV for helium [3, 5], much lower than the energy of a EUV photon, thus eliminating the problem of electron blurring.

The first atom lithography experiment [2] used a proximity lithography setup with a transmission grating placed above a substrate coated with a self-assembled monolayer of alkanethiolates on gold and exposed to a beam of metastable helium atoms. This created a one-to-one image of the transmission grating in the resist. Experiments on pattern generation by manipulating atomic beams by light or electrostatic fields followed [6–9]. However, these methods cannot be used to generate complex patterns with high resolution. For high-resolution pattern generation, experiments have been done focusing atom beams with lenses [10–19], which were then used for serial writing of arbitrary patterns. However, for mass-scale production, serial writing is not a suitable method.

A challenge for pattern generations with metastable atoms is that they decay when impinging on a surface, so using masks made on substrates, as in photolithography, is impossible. The pattern must be generated by open areas in the substrate where the beam can go through. This means any pattern with a closed path (i.e. a circle) would lead to a mask segment falling out.

In 1996, Fujita *et al* [20] came up with an idea to circumvent this problem. They demonstrated pattern generation with metastable neon atoms using a solid mask consisting of a distribution of uniformly sized holes (30 nm in diameter) etched into a silicon nitride membrane. The hole distribution was calculated using the theory of grid-based binary holography developed by Lohmann and Paris [21], and later by Onoe and Kaneko [22] for scalar waves: this theory imposes the limitation that the openings are all of the same sizes and positioned on a regular grid structure. The hole distribution approximates the Fourier transform of the final, desired pattern. In recent publications, binary lithography was further explored, investigating how many holes are needed to generate a particular pattern and what maximum resolution can be achieved with typical matter-wave wavelength scales [9, 23]. However, it was recently shown that the wavefront of metastable helium atoms is perturbed by dispersion forces when it passes through a hole in a silicon nitride membrane [24], an effect not considered in previous publications. Dispersion forces, in this case, the Casimir–Polder forces, are caused by the quantum-mechanical ground-state fluctuations of the electromagnetic field in the absence of charges [25, 26]. Due to the field fluctuations, the helium atom will be polarised for a short time. The resulting induced dipole moment interacts with the dielectric membrane via dipole–dipole interactions. These forces decay dramatically with the distance, r^{-3}-power law [27], but play a significant role on the nanometre length scale and have been observed in several matter-wave diffraction experiments [12, 28–30]. Independent of the atom lithography work, several experiments have been carried out on manipulating atomic and molecular beams using light fields or solid-state diffraction gratings. For example, ground-state atoms (rubidium, Rb [31]; caesium, Cs [32]; sodium, Na [33]; potassium, K [34]; helium, He [35]; neon, Ne [35]; argon, Ar [35]; krypton, Kr [35]; iridium, Ir [30]), metastable atoms (argon, Ar [36]; helium, He [13]), and ground-state molecules (deuterium dimer, D_2 [35]; buckyballs, C_{60} [28] and C_70 [37]; tetraphenylporphyrin [37]; fluorofullerene, $C_{60}F_{48}$ [37]; phthalocyanine, H_2Pc [29]). Several of these experiments are published with theoretical calculations showing the impact of dispersion force interactions in the shape of the diffraction patterns; see, for instance, [28–30, 35, 38].

For Fujita *et al*, the mask holes were so big that the dispersion force interaction was negligible. Hence, they could use the standard binary holography theory to generate the mask design for their patterns. However, the effect of the dispersion force interaction becomes significant for smaller mask holes, which are required for high resolution. [24] shows that metastable helium atoms cannot penetrate holes in 5 nm thick silicon nitride membranes with less than 2 nm diameter.

This paper aims to establish a theoretical framework for mask-based matter-wave lithography, initially in one dimension, to generate a mask for a given desired pattern, atom wavelength, and dispersion interaction. Mathematically this is a so-called inverse problem [39]. We now explain why machine learning is the only realistic approach to solving this problem, even in the one-dimensional case:

As discussed above, the existing binary holography theory for scalar waves uses a grid of holes of equal size as the base for the masks. The work of Born and Wolf furthermore assumes system dimensions and wavelengths that fulfil the well-known condition of optics: Fraunhofer approximation [40]. In the Fraunhofer approximation, the inverse problem is reduced to a Fourier transform of the desired pattern, which yields the required mask. The mask needs to be sampled according to the Nyquist–Shannon sampling theorem to account for the finite size of the holes and mask. In one dimension, this can be done analytically; see section 2.2.1. However, this solution cannot be adapted to matter waves because the dispersion interaction between the particles and the diffraction object induces a complex phase distribution of the matter wave [19].

An alternative approach to Fourier transformation starts from the convolution of the mask function with the so-called point-spread functions [41] representing a basis of elementary openings. This approach, however, raises several issues: (i) the complex structure of the required integral kernel (analogously to the Airy disc for matter waves [38]), which (ii) does not scale linearly with the dimension of the holes [24] and (iii) would require additional optimisation algorithms to consider a variation of the hole sizes.

In this paper, we use machine learning as a new approach to mask generation for metastable atoms; Like Fujita *et al* we remain in the Fraunhofer regime, but we include the dispersion interaction and open up the possibility that holes can be of different sizes and not positioned on a regular grid structure. Our machine-learning architecture is based on a combination of deep learning and genetic algorithms. The problem is solved in two steps: first, a deep neural network [42] is trained on large amounts of generated data. Second, the neural network resolves the desired diffraction pattern, producing a mask. This mask is then mutated to produce an initial population fed to a genetic algorithm [43]. The genetic algorithm then can quickly converge to a mask that produces the desired diffraction pattern with arbitrary precision (see figure 1).

Deep learning has been used extensively to solve inverse problems because neural networks are universal function approximators [44]. For example, it has been applied to light inverse scattering problems [45]. The high-level combination used in this paper (deep learning plus genetics algorithm) has been used in photonic device design [46], inverse molecular design [47], and robot manipulators [48], among others. However, to our knowledge, it has never been used to solve an inverse diffraction problem, where a diffraction mask is recovered from a diffraction pattern.

2. Results

2.1. Mask generation framework

The general idea of our mask generation framework follows: first, a deep convolutional neural network is used to approximate a general solution to the inverse problem. This neural network is trained with an extensive data set of randomly generated examples that map a mask to a diffraction pattern represented by a sequence of real numbers (see section 2.2.3 for details of how the examples are generated). Masks are represented as binary sequences of finite length where 1 represents openings in the mask and 0 blocked space. In the experiments presented here, we chose masks of length 50. Lists of 1000 numbers represent the diffraction patterns.

Once the approximate solution is obtained, this solution (a binary mask) is randomly mutated. The random mutations of this approximate solution form the initial population of a genetic algorithm. The genetic algorithm—known to be an efficient optimiser for binary sequences [49] — further refines the mask provided by the neural network until a convergence constraint is satisfied.

2.1.1. LACENET: a neural network for approximate inversion of the Kirchhoff diffraction formula

LACENET is a deep convolutional neural network that takes as features (inputs) the desired intensity pattern and its fast Fourier transform and has to output the mask pattern, a binary intensity map. We choose the real-input fast Fourier transform as one of the network's features because the Fourier transform of the diffraction pattern plays a significant role in far-field diffraction [9, 40]. Thus, the diffraction patterns have periodic components that can be efficiently compressed by using its Fourier decomposition, and for this reason, allowing for better learning by our neural network.
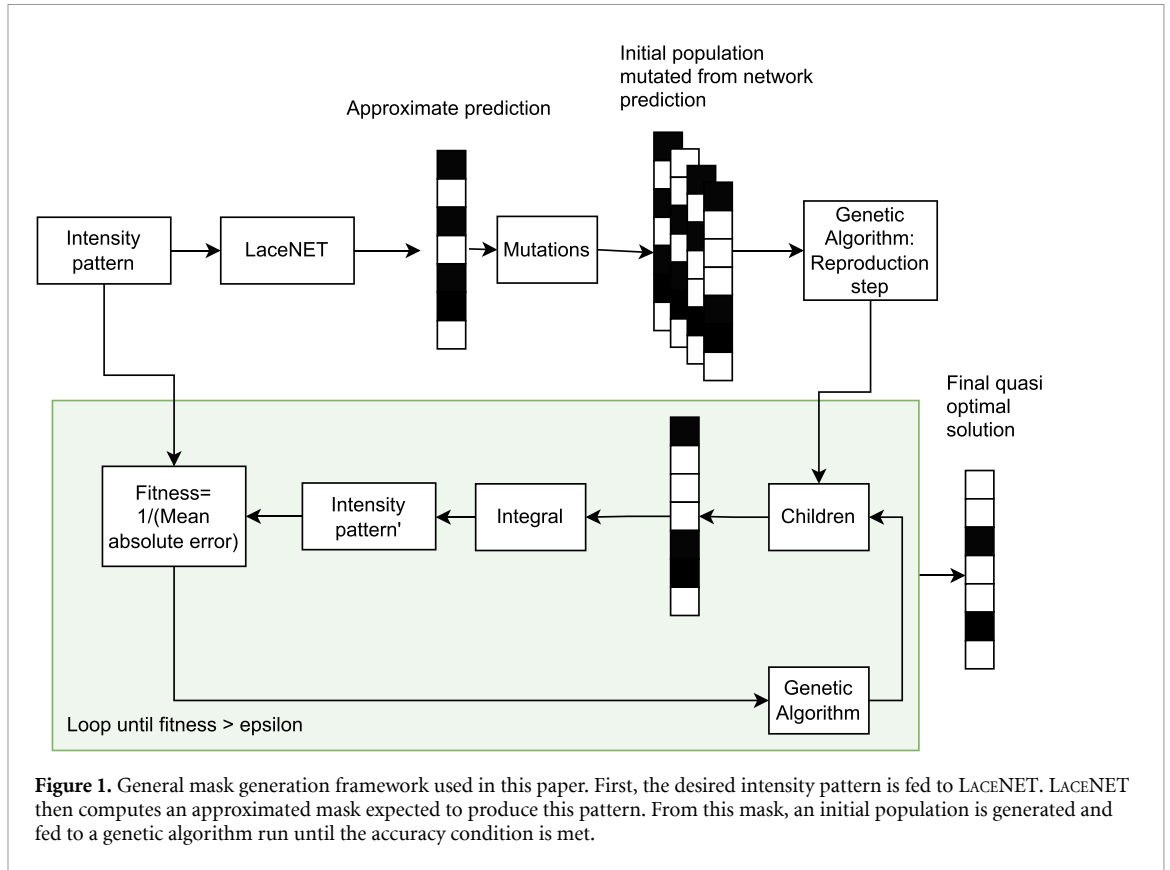
We split the network graph into three towers; each is specialised in one of the three different features (the intensity map and the real and intensity parts of the fast Fourier transform). Within each tower, we use skip connections—a mechanism used effectively in the U-Net architecture, an inverse architecture for light scattering problems [50]. Skip connections consist of taking the input to a neural network layer (or succession of layers) and summing it after the neural network layer has been applied to it [51].

Finally, we use a fully connected layer to learn the predicted mask (see figure 2 for the full architecture). Unlike many popular computer vision architectures, we do not use batch normalisation because we did not see any benefit for our task after extensive experimentation. We train the neural network with focal loss. A loss commonly used to train segmentation masks in computer vision [52]. The focal loss is derived from the binary cross entropy loss

$$l_{\mathrm{BC}}(y,\hat{y}) = [y\log\hat{y} + (1-y)\log(1-\hat{y})] . \tag{1}$$

By summing over all the mask points with the total number N_{points}, we reach the total loss per mask

$$L(y,\hat{y}) = -\frac{1}{N_{\mathrm{pixels}}}\sum_{\mathrm{mask}} l_{\mathrm{BC}}(y,\hat{y}), \tag{2}$$

Figure 1. General mask generation framework used in this paper. First, the desired intensity pattern is fed to LaceNET. LaceNET then computes an approximated mask expected to produce this pattern. From this mask, an initial population is generated and fed to a genetic algorithm run until the accuracy condition is met.

with the true value of each pixel in the mask sequence y and the value predicted by the neural network $\hat{y}$.

The focal loss is a variation of the binary cross-entropy loss and reads

$$l_{\mathrm{FO}}^{i}(y,\hat{y}) = -\alpha_i \left[1 - e^{l_{\mathrm{BC}}(y,\hat{y})} \right]^{\gamma} l_{\mathrm{BC}}(y,\hat{y}). \tag{3}$$
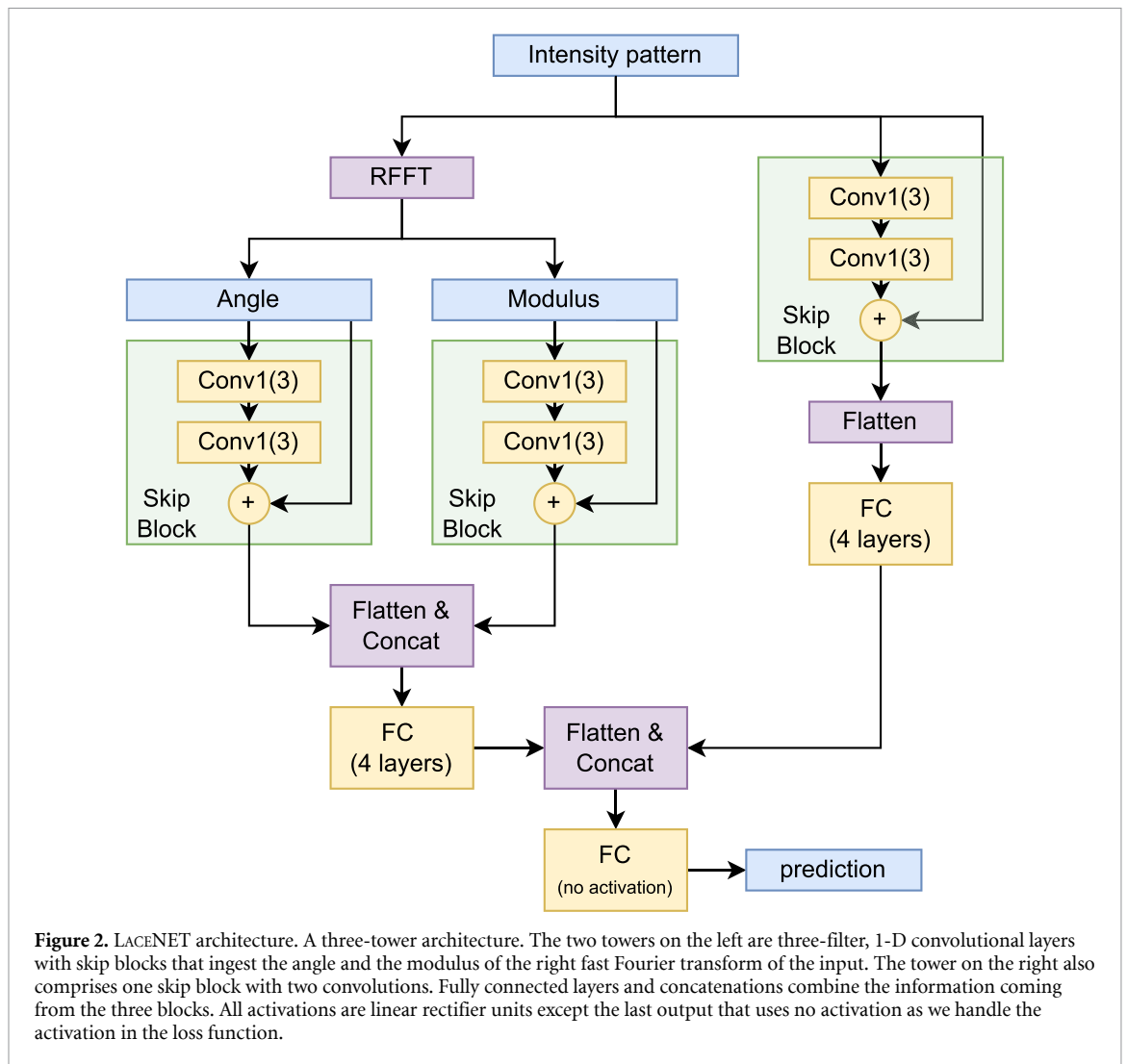
This loss function is designed to achieve two purposes: (I) weight the classes (in our case, 1 and 0) according to their rarity through the weights α_i so that they contribute equally to the loss, and (II) reduce the loss of easily-classified examples, so that the network can focus on the more complex parts of the mask.

LaceNET can be defined as a deterministic non-linear function that depends on its parameters w. The optimisation problem that LaceNET solves as training is to find the set of parameters w that minimise the loss function over the training set Ψ,

$$w^* = \min_{w} \left[\sum_{\Psi} L(y,\hat{y}) \right]. \tag{4}$$

LaceNET is trained on a data set $\mathcal{D}$ of 300k samples, randomly generated as described in section 2.2.3. As it is standard in deep learning, this data set is split into three smaller subsets: a training set $\mathcal{D}_{\mathrm{train}}$, a test set $\mathcal{D}_{\mathrm{test}}$, and a validation set $\mathcal{D}_{\mathrm{val}}$. The split used is 81/9/10 (train/validation/test). One-tenth of the data is reserved for testing, and one-tenth of the remaining training data is reserved for validating hyper-parameters. The training of LaceNET is split into two parts: hyperparameter tuning (which optimises for hyperparameters of the neural network such as learning rate, loss function type etc) and the training of the network itself (i.e. the training of its weights and biases).

We use a variety of Bayesian searches to train the hyperparameters of the neural network and evaluate the validation data set using mean squared error over the result of the integrals of the produced mask. We consider using mean squared error because that is the metric that tells whether the mask produces the required pattern. In the discretised space, the inverse may not be unique and slight differences in the binary sequence that represents the mask might substantially affect the focal loss while affecting the mean squared error loss of the integrated mask to a much lesser extent.

Figure 2. LACENET architecture. A three-tower architecture. The two towers on the left are three-filter, 1-D convolutional layers with skip blocks that ingest the angle and the modulus of the right fast Fourier transform of the input. The tower on the right also comprises one skip block with two convolutions. Fully connected layers and concatenations combine the information coming from the three blocks. All activations are linear rectifier units except the last output that uses no activation as we handle the activation in the loss function.
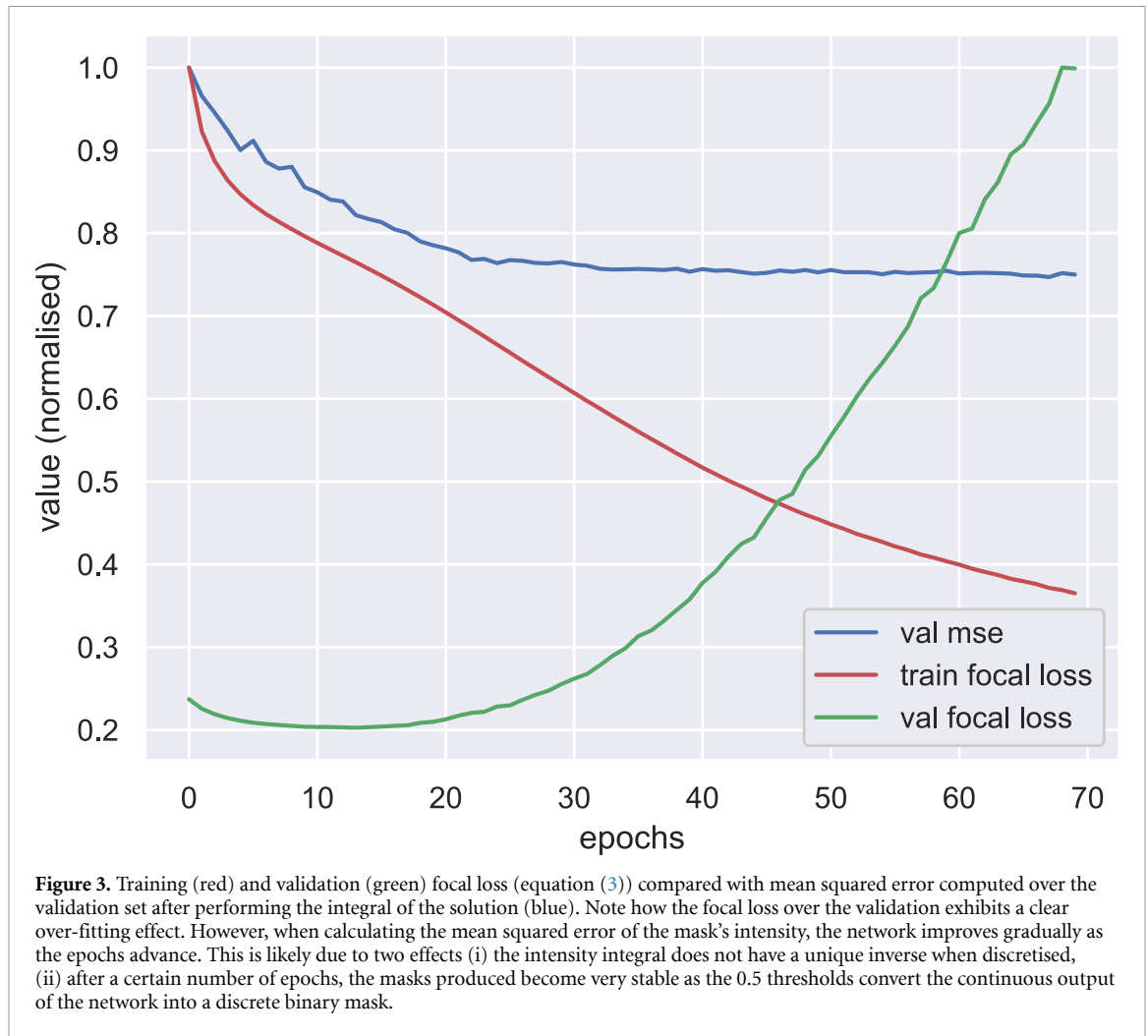
For the training of the neural network, we use the best optimisation algorithm (typically Adam [53], or RMSProp [54]) that is returned by the hyperparameter search. Training is stopped using early stopping with patience 10 and delta 0.001. Full details of this procedure and the space over which hyperparameter search is performed can be found in table A1. The hyperparameters used to produce the results presented here are optimiser: Adam, learning rate: 0.00 016, batch size: 225, α: 0.439, γ: 5.952. Figure 3 shows the focal loss and the mean squared error after integrating the masks evaluated over the validation data set.

2.1.2. Genetic algorithm

Genetic algorithms are stochastic classical evolutionary algorithms, that is, algorithms that dynamically change to optimise a fitness function F. Genetic algorithms are inspired by Darwin's theory of evolution which describes how the genes in the population evolve according to their capacity to reproduce and mutate. In genetic algorithms, the population is not an actual population of living beings but instead a set of solutions for a particular optimisation problem (known as chromosomes). Furthermore, the mutation is not necessarily the result of natural stochastic processes but occurs according to different rules that the programmer can decide. For each chromosome in the population, a fitness value can be calculated using a pre-defined formula set at the researcher's choice. Finally, the reproductive process of life is simulated by combining different chromosomes (parents) into the next generations of offspring. The reproductive process can also be defined by the researcher [55].

Genetic algorithms are very robust optimisation algorithms that are especially suited to work with sequences of categorical variables (such as the four types of bases present in DNA: adenine (A), cytosine (C), guanine (G), and thymine (T)). The masks we want to obtain in this paper are a perfect example: they are entirely defined by a sequence of binary values.

Figure 3. Training (red) and validation (green) focal loss (equation (3)) compared with mean squared error computed over the validation set after performing the integral of the solution (blue). Note how the focal loss over the validation exhibits a clear over-fitting effect. However, when calculating the mean squared error of the mask's intensity, the network improves gradually as the epochs advance. This is likely due to two effects (i) the intensity integral does not have a unique inverse when discretised, (ii) after a certain number of epochs, the masks produced become very stable as the 0.5 thresholds convert the continuous output of the network into a discrete binary mask.
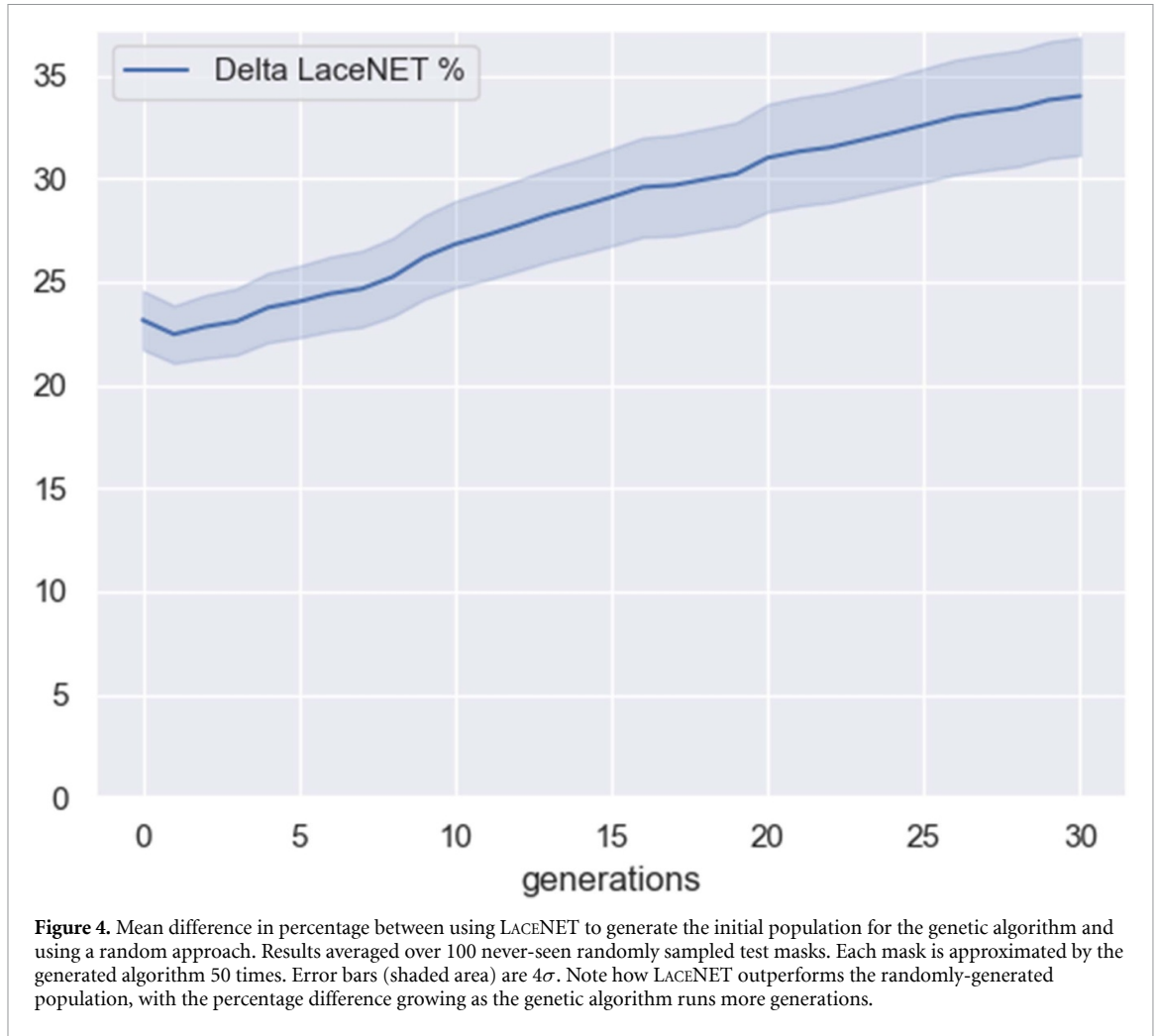
The genetic algorithm that we use here is applied to the solution by LACENET; the fitness function is simply the inverse of the mean absolute error of the generated diffraction pattern plus a small numerical constant α,

$$F = \frac{1}{\alpha + \sum_{r_i} \left| \tilde{P}(r_i) - \psi(r_i; \boldsymbol{x}, \boldsymbol{d}) \psi^{\star}(r_i; \boldsymbol{x}, \boldsymbol{d}) \right|}, \tag{5}$$

where $\tilde{P}(r_i)$ is the absolute square value of the wave function discretised over a grid of radial coordinates r_i measuring the distance to the centre of the diffraction pattern. $\psi(r_i; \boldsymbol{x}, \boldsymbol{d})$ is the wave function that is made to depend on the the positions $\boldsymbol{x} = (x_1, x_2, \dots)$ and thicknesses $\boldsymbol{d} = (d_1, d_2, \dots)$ of the grating (mask) openings. A detailed description of how the wave function is computed can be found in section 2.2. To run our genetic algorithm, we use a pygad: a well-known genetic algorithm solver for python [56]. Within pygad we use the following hyper-parameters (obtained through grid search). (i) uniform crossover, in which random recombining parents' chromosomes form offspring. (ii) fit parent persistence: the fittest chromosomes are carried on to the next generation. (iii) initial population size of 50 chromosomes. (iv): The number of solutions selected as parents are set to 7.

The initial population is not generated by pygad but by our bespoke script. The script inputs the number of mutations allowed in each chromosome and generates a population by randomly mutating the output of LACENET. LACENET's solution is also kept as part of the initial population. The results presented here have 15 mutations in the chromosome. Mutations are assigned sampling randomly from a discrete uniform distribution—which means that gen randomly set to mutate can maintain its initial value if it is sampled again from the distribution.

Figure 4 shows the average difference in percentage when using LACENET to generate the initial solution population and using a random initial population. Note that within 4σ error bars, LACENET outperforms a naive genetic approach. More importantly, perhaps, LACENET becomes better than using a randomly

Figure 4. Mean difference in percentage between using LaceNET to generate the initial population for the genetic algorithm and using a random approach. Results averaged over 100 never-seen randomly sampled test masks. Each mask is approximated by the generated algorithm 50 times. Error bars (shaded area) are 4σ. Note how LaceNET outperforms the randomly-generated population, with the percentage difference growing as the genetic algorithm runs more generations.

initiated genetic algorithm the more generations the genetic algorithm is run, showing that the edge provided by the neural network increases with computational time.
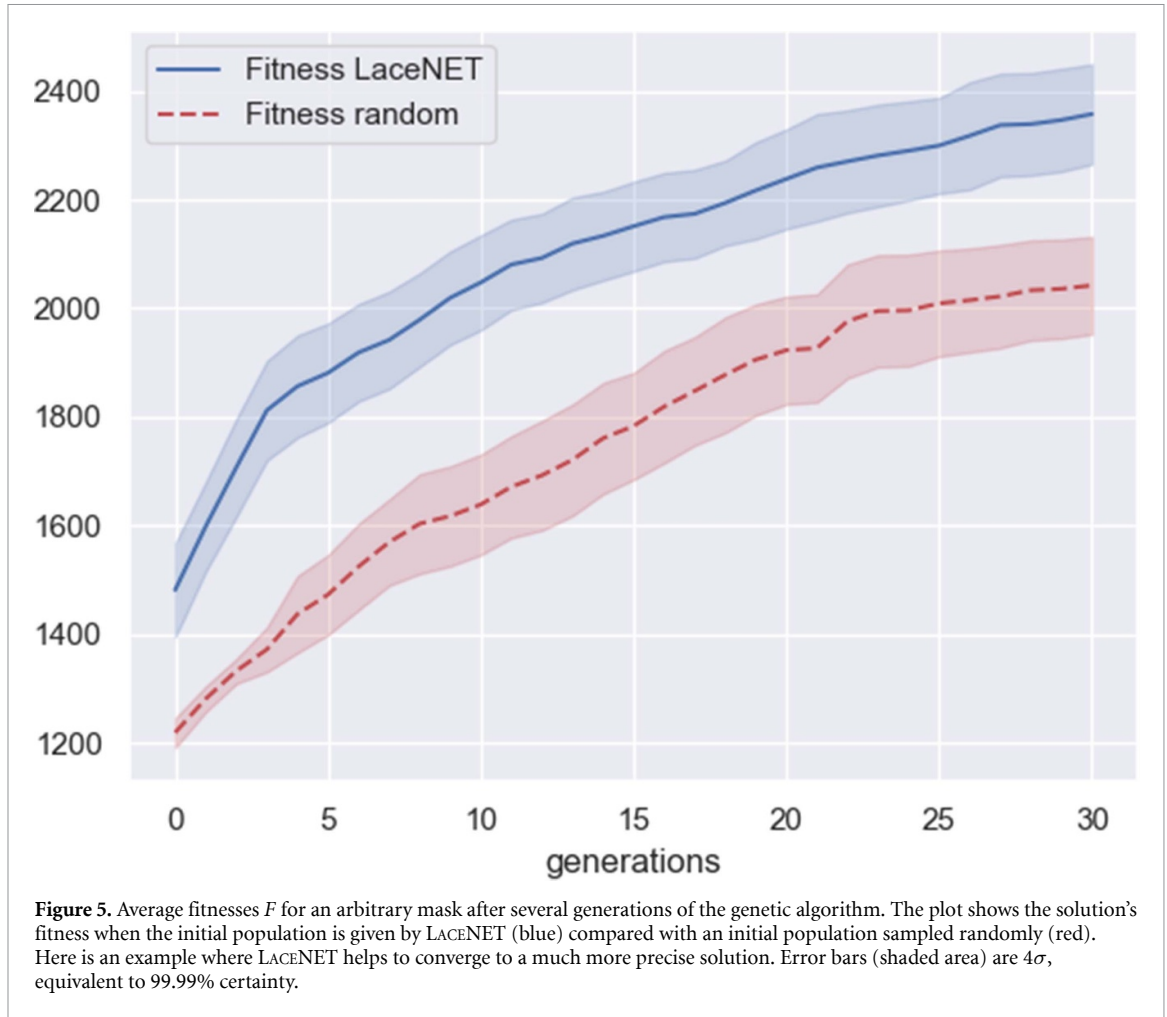
Figure 5 shows an example of a particular mask in which LaceNET performs significantly better than random. Results for all 100 randomly generated masks can be found in appendix B. As shown above, when all masks are averaged, LaceNET outperforms random significantly (by more 4σ). Figure 6 compares the solutions returned by LaceNET and the genetic algorithm for a particular inverse problem. Note how LaceNET provides a reasonable first approximation to mask design and how further Genetic iteration incorporates small changes (mutations) that make the solution closer to the ground truth.

2.2. Forward propagation

Here we describe how the forward propagation of matter waves generates the training patterns through randomly generated masks. The diffraction at a dielectric interface can be described via classical waves due to the duality of waves and particle [57], which can be derived via Kirchhoff's diffraction formula [40]. This method is commonly applied to describe the interference of matter waves in eikonal (or paraxial) approximation [40], which is valid when the transversal momentum of the matter wave is negligible compared to the longitudinal ones. [28–30, 38] This formula determines the propagation of a classical wave with amplitude a_0

$$\psi_{\mathrm{cl}}(\boldsymbol{r}_{\mathrm{D}}) = \frac{a_0 k_0}{2\pi\,\mathrm{i}} \int \mathrm{d}^2 r_{\mathrm{i}}\, T(\boldsymbol{r}_{\mathrm{i}}) \frac{\mathrm{e}^{\mathrm{i}k_0(r_{\mathrm{S,i}}+r_{\mathrm{i,D}})}}{r_{\mathrm{S,i}} r_{\mathrm{i,D}}} \frac{\cos\vartheta + \cos\vartheta'}{2}, \tag{6}$$

through an interface, in analogy to Huygens–Fresnel principle [58]. The interface is usually described by a transmission function T, which varies (for electromagnetic waves) between 1 for transmission and 0 for absorption. The source of the wave is located at $\boldsymbol{r}_{\mathrm{S}}$ (subscript S for source), and the grating is placed at the intermediate points $\boldsymbol{r}_{\mathrm{i}}$ (subscript i for intermediate), leading to the relative coordinates $\boldsymbol{r}_{\mathrm{S,i}} = \boldsymbol{r}_{\mathrm{S}} - \boldsymbol{r}_{\mathrm{i}}$ to the grating and continues the propagation to the detector $\boldsymbol{r}_{\mathrm{D}}$ (subscript D for detector) $\boldsymbol{r}_{\mathrm{i,D}} = \boldsymbol{r}_{\mathrm{i}} - \boldsymbol{r}_{\mathrm{D}}$. The wave

Figure 5. Average fitnesses F for an arbitrary mask after several generations of the genetic algorithm. The plot shows the solution's fitness when the initial population is given by LaceNET (blue) compared with an initial population sampled randomly (red). Here is an example where LaceNET helps to converge to a much more precise solution. Error bars (shaded area) are 4σ, equivalent to 99.99% certainty.

vector is related via the de-Broglie wavelength $k_0 = 2\pi/\lambda_{\text{dB}} = p/\hbar$ with the particle's momentum $\boldsymbol{p} = p\boldsymbol{e}_p$. [57] The geometric correction angles ϑ and ϑ' are the angles between the aperture's normal $\boldsymbol{n}$ and the wave's propagation directions, $\boldsymbol{r}_{\text{S,i}}$ and $\boldsymbol{r}_{\text{i,D}}$, respectively.

By considering plane waves passing the obstacle, the propagation lengths are dominated by the distances between the source and the interface $r_{\text{S,i}} \approx L_1$ and between the interface and the detector $r_{\text{i,D}} \approx L_2$. This approach simplifies Kirchhoff's diffraction formula (6) to the Fourier transform of the transmission function and is known as Fraunhofer approximation

$$\psi_{\text{cl}}(r) = \frac{a_0 k_0}{2\pi i} \frac{e^{ik_0\left[L_1+L_2+r^2/(2L_2)\right]}}{L_1 L_2} \int dA\, T e^{i\frac{k_0}{L_2}\boldsymbol{r}\cdot\boldsymbol{s}}, \tag{7}$$
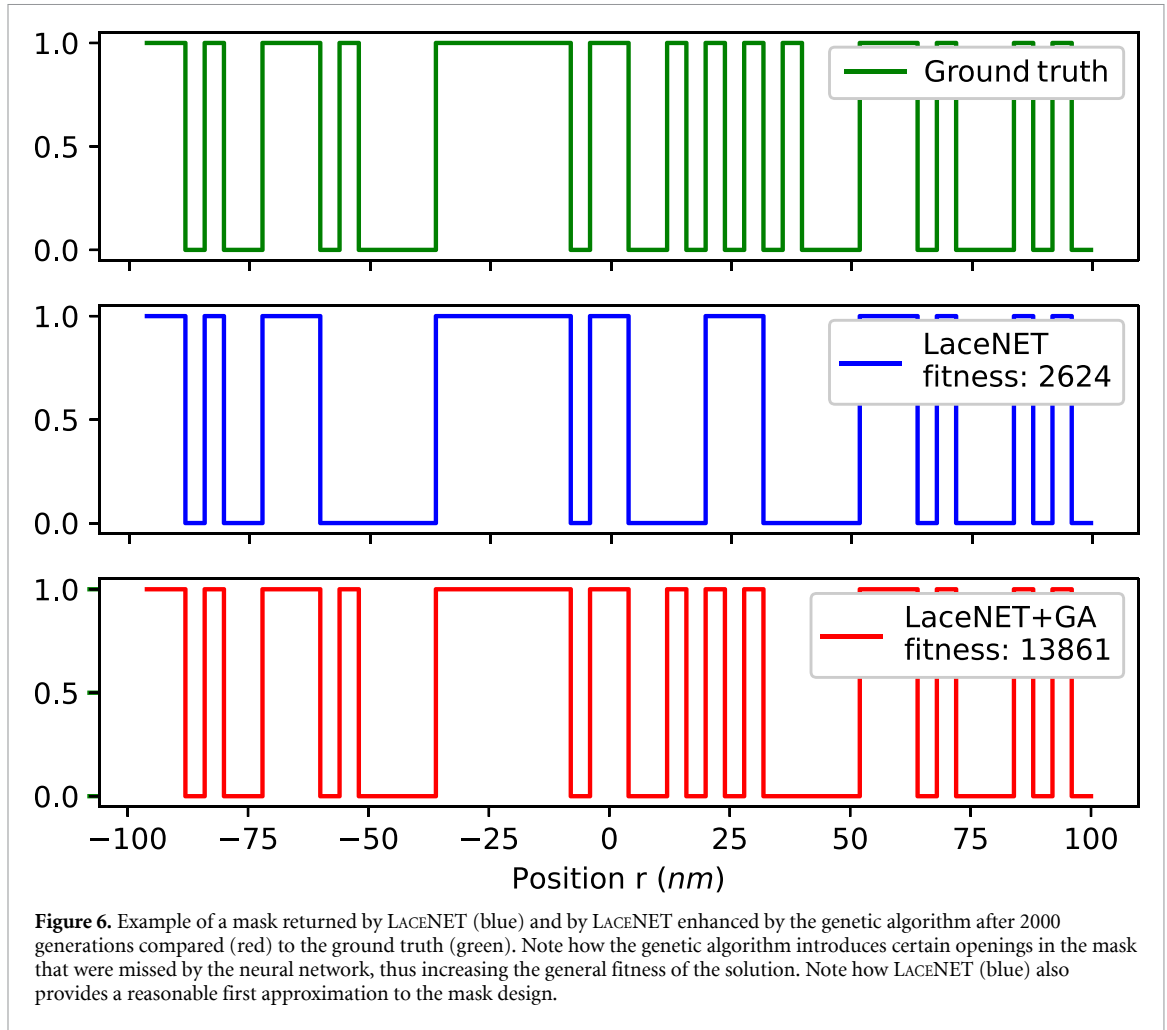
where r denotes the position on the screen. Furthermore, the transmission function is invariant along the direction of the slit, and thus, the wave at the screen is just given by the one-dimensional Fourier transform of the transmission function

$$\psi_{\text{cl}}(r) = \frac{a_0 k_0}{2\pi i} \frac{e^{ik_0\left[L_1+L_2+r^2/(2L_2)\right]}}{L_1 L_2} \int dx\, T(x) e^{i\frac{k_0}{L_2}rx}. \tag{8}$$

Assuming the mask to be an array of slit openings, the superposition of the single slits can determine the total transmission function

$$T(x) = \sum_n t(x - x_n; d_n), \tag{9}$$

with the centre coordinate of each opening x_n. We allow each opening to vary in its width d_n. By inserting the transmission function (9) into the wave equation (8), the field at the screen decomposes into the

Figure 6. Example of a mask returned by LACENET (blue) and by LACENET enhanced by the genetic algorithm after 2000 generations compared (red) to the ground truth (green). Note how the genetic algorithm introduces certain openings in the mask that were missed by the neural network, thus increasing the general fitness of the solution. Note how LACENET (blue) also provides a reasonable first approximation to the mask design.

superposition of single slits with a phase correction due to the corresponding spatial shift

$$\psi(r) = \frac{a_0 k_0}{2\pi i} \frac{e^{ik_0\left[L_1+L_2+r^2/(2L_2)\right]}}{L_1 L_2} \sum_n e^{-i\frac{k_0}{L_2}rx_n} \int dx\, t(x; d_n) e^{i\frac{k_0}{L_2}rx}. \tag{10}$$

The interference pattern is given by the absolute square value of the wave function

$$P(r) = \psi(r)\psi^\star(r). \tag{11}$$

This yields that a target pattern $\tilde{P}(r)$ can be approximated by minimising the error function

$$E(\boldsymbol{x}, \boldsymbol{d}) = \int dr \left(\tilde{P}(r) - \psi(r; \boldsymbol{x}, \boldsymbol{d})\psi^\star(r; \boldsymbol{x}, \boldsymbol{d})\right)^2, \tag{12}$$

with respect to the positions $\boldsymbol{x} = (x_1, x_2, \dots)$ and thicknesses $\boldsymbol{d} = (d_1, d_2, \dots)$ of the grating openings. Ordinarily, one would apply a least-square fitting algorithm to reduce the error (12) concerning the positions and grating openings.

2.2.1. Propagation of electromagnetic waves
In the absence of dispersion forces, which is achieved for electromagnetic waves, the wave propagation (10) can be calculated explicitly, leading to

$$\psi(r) = \frac{a_0}{\pi i} \frac{e^{ik_0\left(L_1+L_2+r^2/(2L_2)\right)}}{L_1 r} \sum_n e^{-i\frac{k_0}{L_2}rx_n} \sin\left(\frac{k_0 r d_n}{2L_2}\right), \tag{13}$$

and, thus, the interference pattern reads

$$P(r) = \frac{a_0^2}{\pi^2 L_1^2 r^2} \sum_{n,m} e^{i\frac{k_0}{L_2}r(x_m-x_n)} \sin\left(\frac{k_0 r d_n}{2L_2}\right) \sin\left(\frac{k_0 r d_m}{2L_2}\right). \tag{14}$$

To analyse the relationship between the interference pattern $P(r)$ and the transmission function parameters $\boldsymbol{x}$ and $\boldsymbol{d}$ concerning the resolution and contrast of the interference pattern, a Fourier analysis is required, which yields [59]

$$
\begin{aligned}
P(\kappa) = -\frac{a_0^2}{2L_1^2} \sum_{n,m} & \left\{ \left(\kappa - \frac{c_1}{2\pi}\right) \left[2\theta\left(\kappa - \frac{c_1}{2\pi}\right) - 1 \right] \right. \\
& + \left(\kappa - \frac{c_2}{2\pi}\right) \left[2\theta\left(\kappa - \frac{c_2}{2\pi}\right) - 1 \right] \\
& \left. + \left(\kappa - \frac{c_3}{2\pi}\right) \left[2\theta\left(\kappa - \frac{c_3}{2\pi}\right) - 1 \right] + \left(\kappa - \frac{c_4}{2\pi}\right) \left[2\theta\left(\kappa - \frac{c_4}{2\pi}\right) - 1 \right] \right\},
\end{aligned}
\tag{15}
$$

with the Heaviside function θ and the frequency shifts

$$
c_1 = \frac{k_0}{2L_2}\left(2x_m - 2x_n - d_n + d_m\right), c_2 = -\frac{k_0}{2L_2}\left(-2x_m - 2x_n - d_n + d_m\right),
\tag{16}
$$

$$
c_3 = \frac{k_0}{2L_2}\left(2x_m - 2x_n + d_n + d_m\right), c_4 = -\frac{k_0}{2L_2}\left(-2x_m - 2x_n + d_n + d_m\right).
\tag{17}
$$

This analysis illustrates that the relevant contributions are located in the range $\kappa = \mathcal{O}\left(10^{-5}/\lambda_{\mathrm{dB}}\right)$ for typical parameters used in matter-wave lithography [spatial resolution of the mask $\Delta x = \mathcal{O}(\mathrm{nm})$ and distance between the mask and the object plane $L_2 = \mathcal{O}(100\,\mu\mathrm{m})$]. By increasing the mask's extension to be in the same order as L_2, the entire spectral resolution will be in the range $10^{-5}/\lambda_{\mathrm{dB}} \le \kappa \ll 1/\lambda_{\mathrm{dB}}$. The Fresnel number determines the upper bound of the resolution to stay in the Fraunhofer regime. The resolution (smallest structure) is determined by the largest mode κ leading to the relation for periodic gratings

$$
\Delta x = \frac{\lambda_{\mathrm{dB}} L_2}{d + D},
\tag{18}
$$

with the opening size d and illumination range D.

However, due to the presence of dispersion forces, applying such a method would result in several numerical issues caused by the complexity of the dependence on these quantities, which will be illuminated in the next section.

2.2.2. Dispersion force interactions

The dispersion force interaction between an atom and a dielectric membrane can be approximated by [24]

$$
U_{\mathrm{CP,app}}(\boldsymbol{r}) = -\frac{9C_3}{\pi} \int \frac{\mathrm{d}^3 s}{|\boldsymbol{s} - \boldsymbol{r}|^6},
\tag{19}
$$

with the C_3-coefficient denoting the interaction strength between the atom and a plane built of the same material as the membrane, the position of the atom $\boldsymbol{r}$, and the integration volume of the membrane's surface bound $\mathrm{d}^3 s$.

The reduction of the slit opening can be found analogously to [24] (table 3) due to the applied assumption of a single-wall interface. However, the phase shift needs to be adapted and reads for a membrane of thickness w with a slit opening d

$$
\varphi(x) \approx -\frac{m\lambda_{\mathrm{dB}}}{2\pi\hbar^2} \int U(x,z)\mathrm{d}z = -\frac{12C_3 m\lambda_{\mathrm{dB}} wd\left(d^2 + 12x^2\right)}{\hbar^2 \pi \left(d + 2x\right)^3 \left(d - 2x\right)^3}.
\tag{20}
$$

To this end, the transmission function for a single slit adapts due to the dispersion force interaction to

$$
t(x) = \theta\left(\frac{d - 2\Delta R}{2} - x\right) \theta\left(\frac{d - 2\Delta R}{2} + x\right) \mathrm{e}^{\mathrm{i}\varphi(x)}.
\tag{21}
$$

The impact of the dispersion forces (Casimir–Polder) is illustrated in figure 7, which shows the difference in the interference pattern from the classical double slit experiment using a scalar wave and a matter-wave with dispersion force interaction. A much higher population of the higher diffraction orders can be observed. A typical helium source generates a flux of 10^{10} atoms/(mm^{2}s) [60] leading to a population of $\approx 800\,\mathrm{atoms}/(\mathrm{mm\,s})$ in the zeroth order.

In older works, see, for instance, [20], the impact of the dispersion forces has been neglected due to the large openings. Due to the r^{-3} power law of the dispersion force near an interface (20), its dominant impact occurs in a region close to the bar's surface up to a distance $\bar{x}$ for the surface. Thus, if this length is much
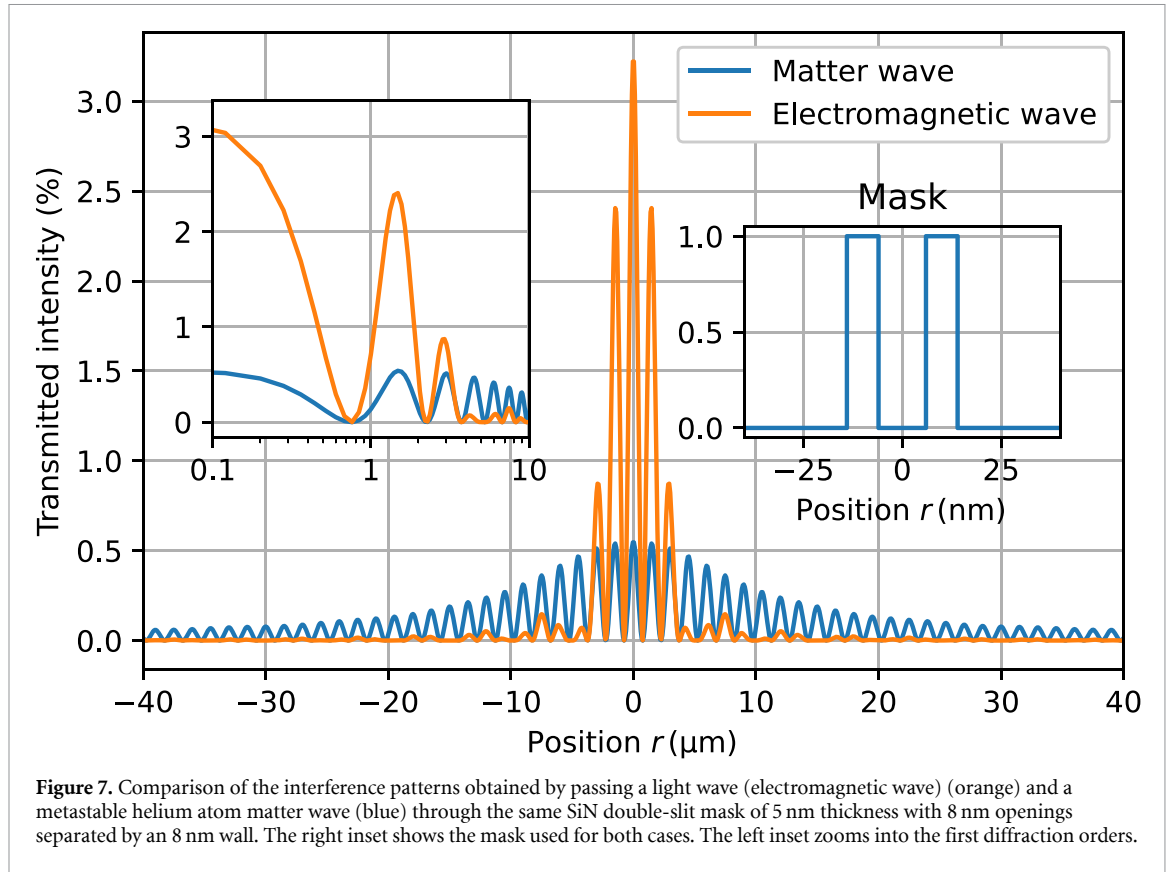
Figure 7. Comparison of the interference patterns obtained by passing a light wave (electromagnetic wave) (orange) and a metastable helium atom matter wave (blue) through the same SiN double-slit mask of 5 nm thickness with 8 nm openings separated by an 8 nm wall. The right inset shows the mask used for both cases. The left inset zooms into the first diffraction orders.

smaller than the opening d, $2\bar{x}/d \ll 1$, the impact of the dispersion force is negligible. To estimate the threshold, we consider the diffraction from a single opening,

$$\psi'(r) = e^{ik_0 L_1} e^{ik_0\left(L_2 + \frac{r^2}{2L_2}\right)} \int\limits_{-d/2}^{d/2} e^{i\frac{C_3 w}{\hbar v_z}\left[\frac{1}{\left(x - \frac{d}{2}\right)^3} - \frac{1}{\left(\frac{d}{2}+x\right)^3}\right]} e^{-i\frac{k_0}{L_2}rx} dx, \qquad (22)$$

where we only insider the dominant part of the dispersion potential inside the opening and neglect the particle's arrival and departure [29]. Furthermore, we neglect the impact of the opening reduction [24]. This field in the target plane has to be compared with the corresponding field of an electromagnetic wave

$$\psi(r) = e^{ik_0 L_1} e^{ik_0\left(L_2 + \frac{r^2}{2L_2}\right)} \int\limits_{-d/2}^{d/2} e^{-i\frac{k_0}{L_2}rx} dx, \qquad (23)$$

to estimate the relative deviation

$$E = \frac{1}{d} \int\limits_{-\infty}^{\infty} \left[\psi(r) - \psi'(r)\right]^2 dr. \qquad (24)$$

By applying Parseval's theorem and substituting $x \mapsto 2x/d$ in the remaining integral, the error can be evaluated numerically by substituting $\tilde{a} = 8C_3 w/(\hbar v_z d^3)$ and a critical value for $\tilde{a}$ can be found by setting the deviation 1% leading to $\tilde{a} = 2.0448 \cdot 10^{-4}$. Finally, we find the critical opening

$$d_C = \frac{2}{\hbar} \sqrt[3]{\frac{C_3 w \hbar^2}{\tilde{a} v_z}} \approx \frac{33.95}{\hbar} \sqrt[3]{\frac{C_3 w \hbar^2}{v_z}}. \qquad (25)$$

If the opening is larger than this critical value, $d > d_C$, the impact of the dispersion force can be neglected. In our case, the critical opening is larger than 100 nm.

This critical opening d_C describes the upper bound for the impact of the dispersion forces. Its lower bounds are determined by the reduction of the grating opening ΔR, considered in previous works [24],

where we analysed the particle's trajectories terminating in the grating bar yielding a minimum distance to it for the particle to enter. Due to the high velocity of the particles meaning short interaction times, the classical consideration of the trajectories is sufficient. This reduction is around 1 nm for our case. Finally, the considered openings need to be in the range

$$2\Delta R < d < d_C, \tag{26}$$

to observe the dispersion interaction's impact and transmit the matter wave.

2.2.3. Generation of training data

In this proof of concept paper, we have considered a 1D system with dimensions so that the Fraunhofer approximation applies. We use a typical wavelength for metastable helium $\lambda_{dB} = 0.1$ nm (corresponding to a particle velocity of $\approx 1000 \, \text{m s}^{-1}$), assuming a helium point-source located $L_1 = 1$ m away from the mask, and a substrate patterning plane at $L_2 = 300 \, \mu$m behind the mask. The mask is considered to be made of SiN with a thickness of 5 nm. Such a mask will stay stable under the penetration with Helium atoms due to their small kinetic energy of 20–25 meV. In the experiment reported in [29], the authors used grating with doubled thickness, and the molecules had a kinetic energy of up to 650 meV. The maximum extension of the mask is restricted to $D = 200$ nm to stay within the Fraunhofer approximation (satisfying a Fresnel number below 1). Furthermore, the mask is separated into 50 sections of 4 nm width each, corresponding to the minimum opening width. This width was selected based on [24], which shows that the effective entrance width is reduced by about 2 nm due to the dispersion force interaction.

We randomly generate masks represented via arrays with entries 0 for close or 1 for open sections. Neighbouring open sections will be combined into a single opening with a larger thickness. Hence, the randomly generated array will be transferred into a set of widths and positions for each opening. Thus, the allowed openings vary between 1 and 50 units meaning the network is trained on single-slit diffraction and multi-slit interference with up to 25 openings leading to an angular dispersion of $1 - 7 \, \text{deg nm}$. To this end, we calculate the single slit diffraction (integrant in equation (10))

$$\int dx t(x; d_n) e^{i \frac{k_0}{L_2} r x}, \tag{27}$$

for all possible thicknesses d_n for matter waves (21) and electromagnetic waves (13) via standard numerical integration techniques. These results are tabled, and a second program calculates multi-slits' superposition to generate the training data.

2.3. An application example: the double slit pattern

To demonstrate the performance of LACENET, we invert a well-known diffraction pattern: the pattern resulting from a double slit mask, see figure 7. Note that a double slit differs greatly from the LACENET's training set, consisting of randomly generated masks with no particular preset structure.
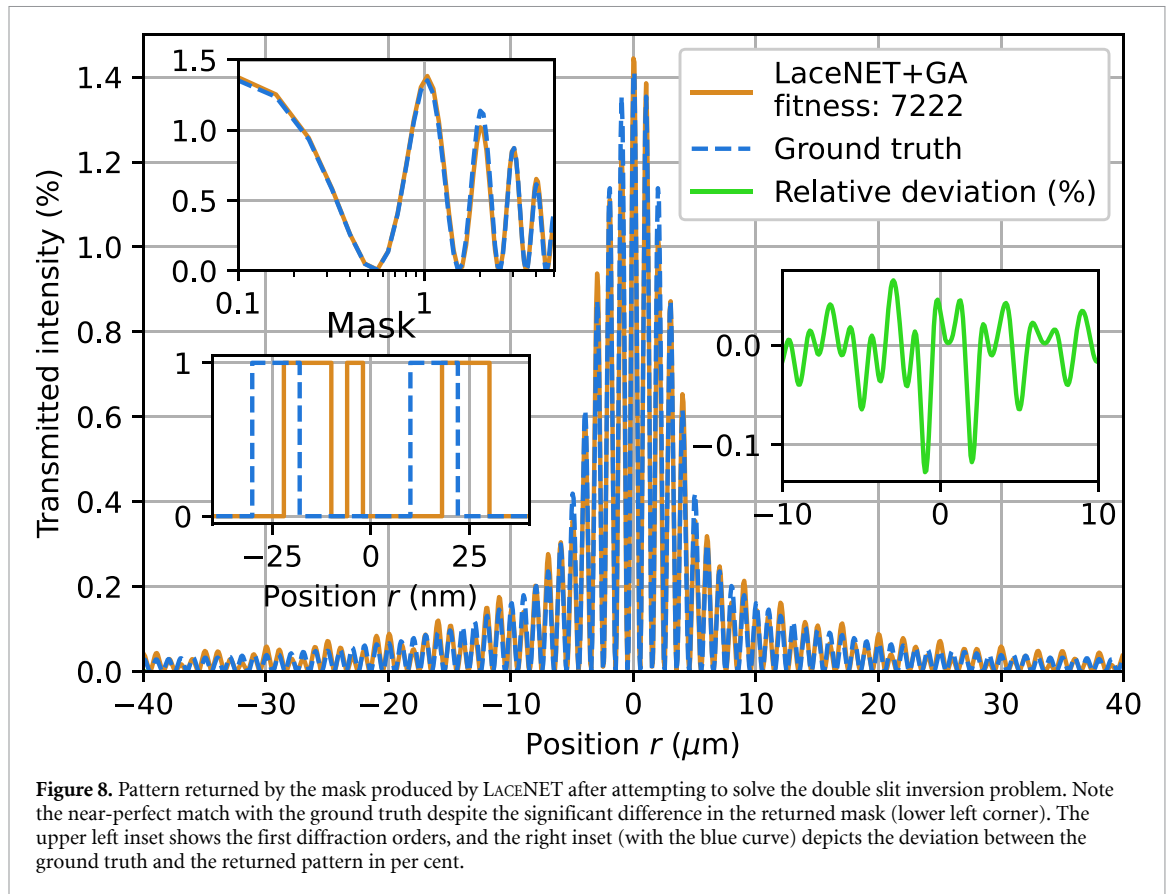
LACENET successfully inverts the pattern to a good degree of accuracy after 10 000 generations of the genetic algorithm. Interestingly, it does not obtain the same mask as the ground truth, but the mask it produces matches very closely to the desired pattern (see figure 8).

3. Discussion

Despite the known capabilities of deep learning as a powerful inverse strategy, we found that the best method to obtain accurate results was to couple it with an additional optimisation step (in our case, a genetic algorithm). LACENET's difficulty in fully solving the inverse could be due to (i) insufficient data, (ii) low network complexity, or (iii) the challenging inverse problem that it is trying to solve.

We performed several experiments with some examples to determine the cause of LACENET's difficulty. Interestingly, increasing the complexity of the network (by adding layers and increasing the number of neurons in the layers) did not produce a better generalisation. With regards to point (ii), we did see a clear improvement with increasing the size of our data set (which we can arbitrarily scale up by solving for equation (27)). The experiments presented here use a data set of 300k examples. Using the data generation procedure described in section 2.2.3, one can extend the data set arbitrarily, potentially increasing the generalisation capabilities of LACENET.

Finally, we saw some examples where initialising the mask using a random sequence or using LACENET made no significant difference. By observing these examples, it seems to be the case that this happens more often for masks that contain many small openings. These masks are more likely to be automatically randomly produced by the genetic algorithm. Therefore, it is possible that for some cases, the random initialisation is as good as, or even better than LACENET's guess.

Figure 8. Pattern returned by the mask produced by LACENET after attempting to solve the double slit inversion problem. Note the near-perfect match with the ground truth despite the significant difference in the returned mask (lower left corner). The upper left inset shows the first diffraction orders, and the right inset (with the blue curve) depicts the deviation between the ground truth and the returned pattern in per cent.

4. Conclusion

We present a machine learning architecture for solving the inverse problem for matter-wave lithography in 1D in the Fraunhofer regime using a monochromatic plane wave: LACENET. The architecture of LACENET consists of a convolutional deep neural network followed by a genetic algorithm. This is a key technological step for industrial-scale matter-wave lithography, needed to achieve the ultimate goal of fast patterning individual atoms and molecules over large areas. The dispersion force interaction for matter waves presents a tremendous computational challenge compared to scalar waves due to its complex integration and nonlinear solution.

In this proof of concept paper, we have restricted ourselves to the 1D case in the Fraunhofer approximation regime. Future work will focus on modifying LACENET to solve the general problem in 2D and extending the method to incorporate polychromatic waves and extended sources.

Data availability statement

The data cannot be made publicly available upon publication because they are owned by a third party and the terms of use prevent public distribution. The data that support the findings of this study are available upon reasonable request from the authors.

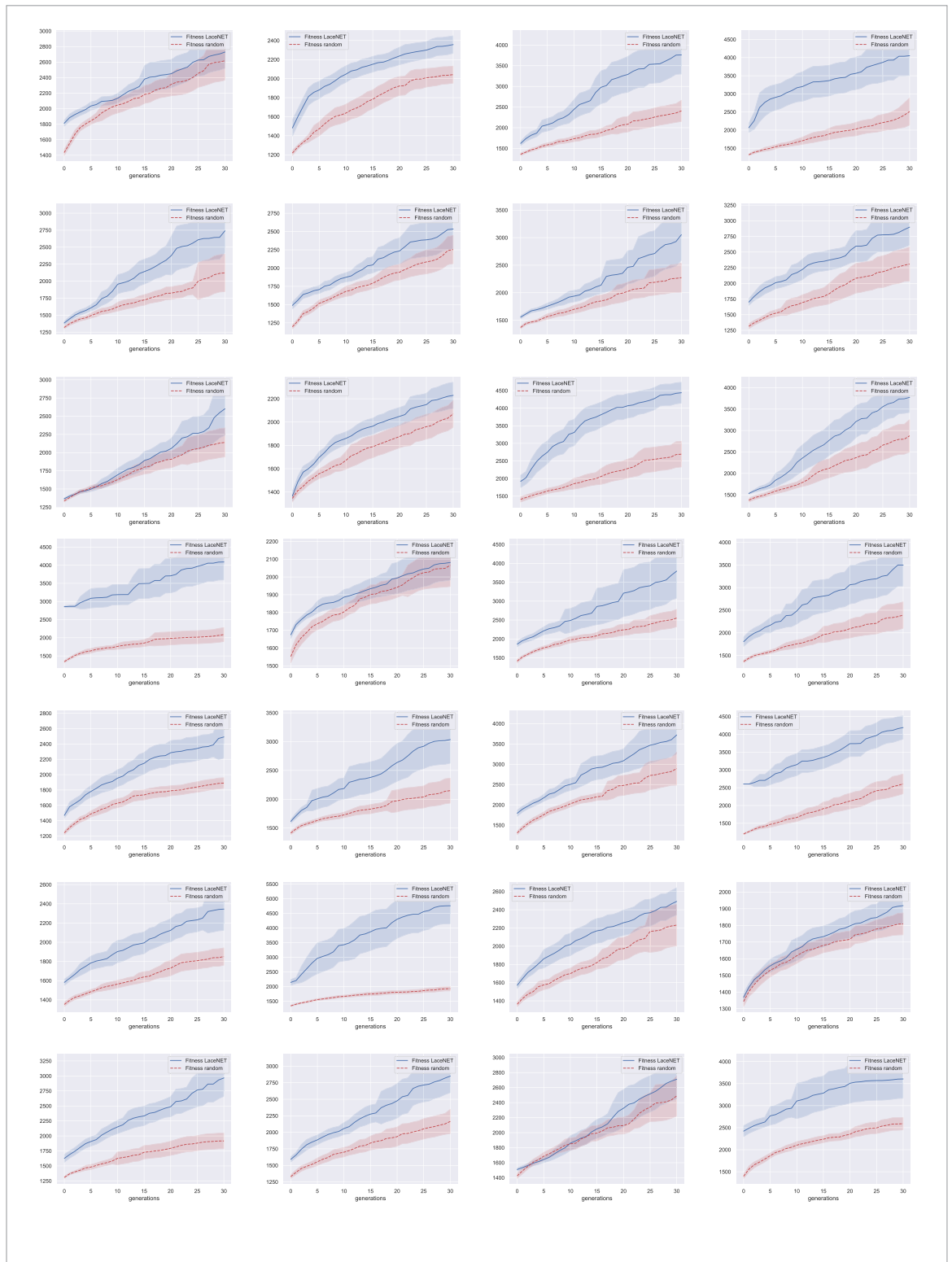### Acknowledgments

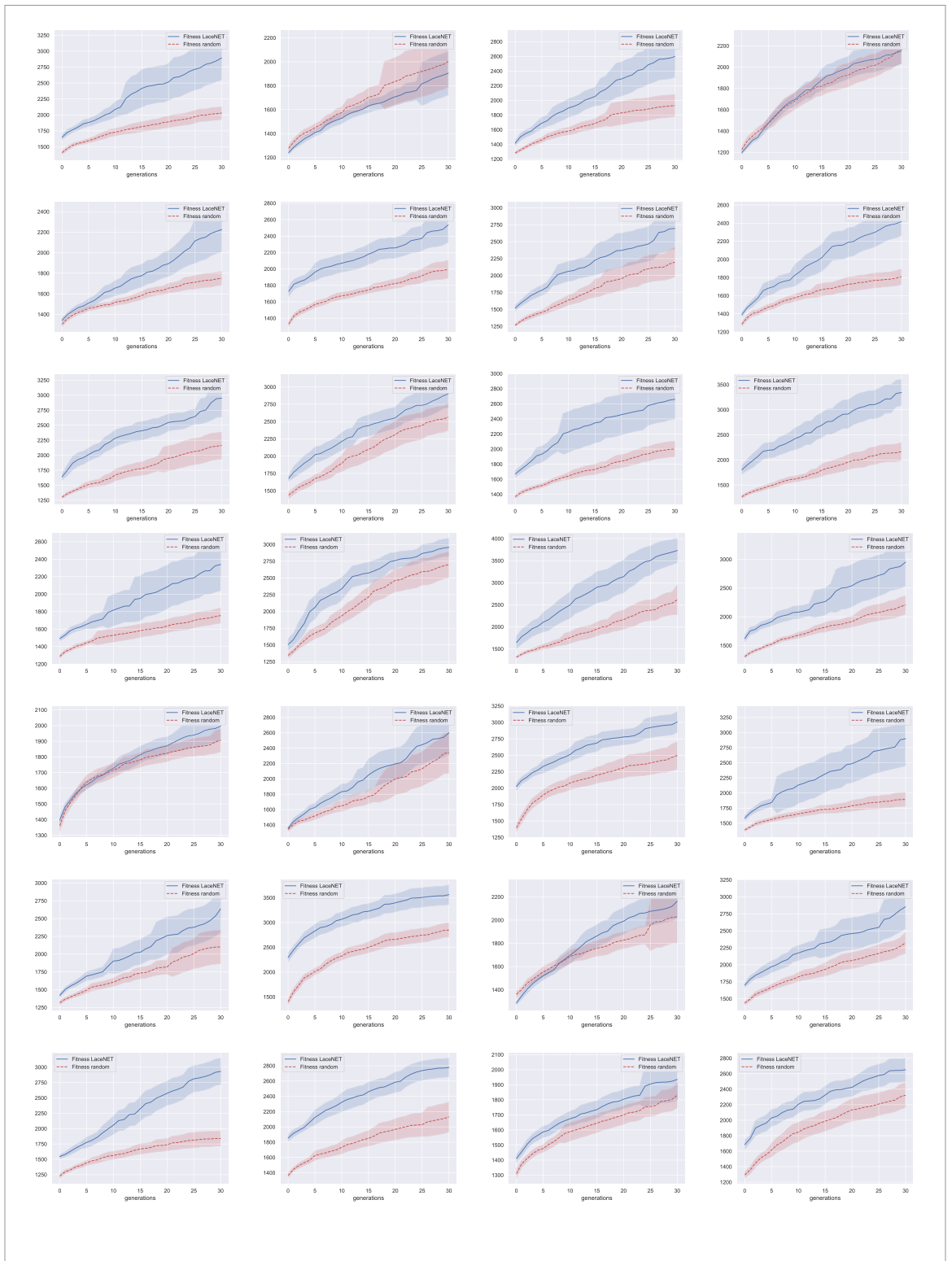## Appendix A. Hyperparameter tuning

Table A1. Results of the hyperparameter tuning rounds in LACENET.
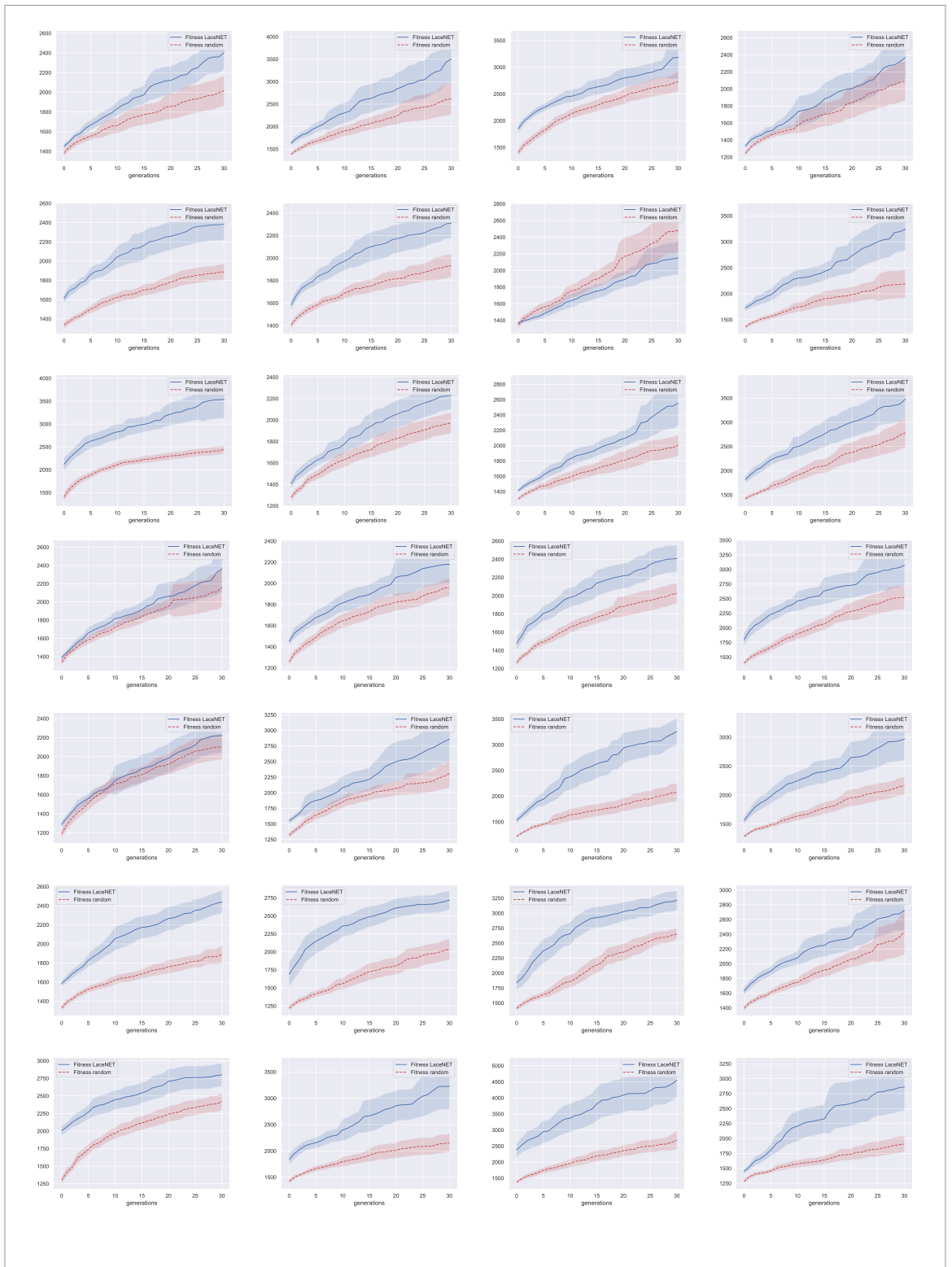
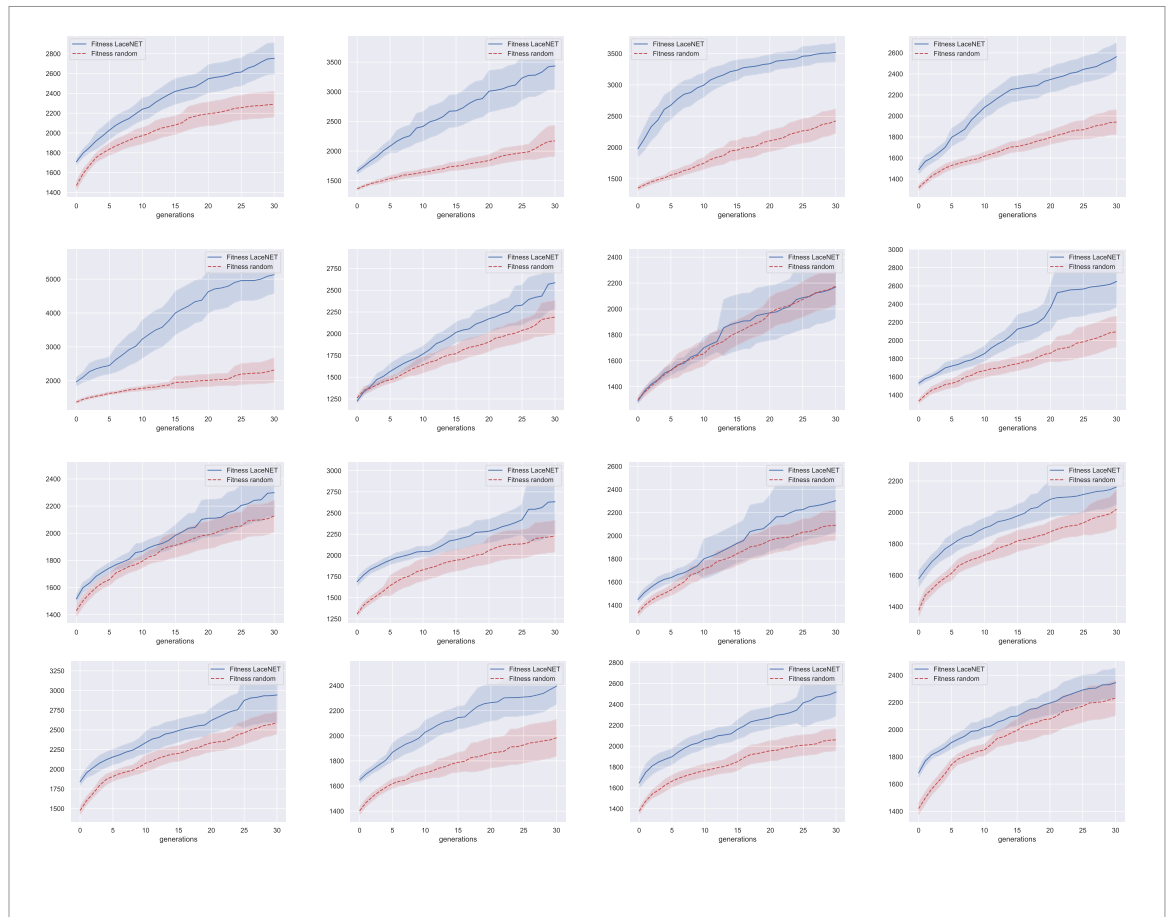|	optimizer	lr	batch_size	alpha	gamma	MSE
11	Adam	0.000 163	225	0.439 672	5.951 962	1.399 777
43	RMSprop	0.000 288	225	0.448 443	6.431 573	1.404 423
47	RMSprop	0.000 123	87	0.439 803	6.291 578	1.421 799
48	RMSprop	0.000 283	200	0.448 358	6.074 838	1.433 143
39	RMSprop	0.000 099	66	0.466 304	7.131 291	1.450 584
4	Adam	0.000 098	225	0.584 123	5.253 094	1.456 233
22	Adam	0.000 173	225	0.40 145	5.910 984	1.456 791
23	Adam	0.000 109	225	0.450 265	5.525 971	1.465 228
12	Adam	0.000 187	225	0.412 308	6.12 187	1.465 743
32	RMSprop	0.000 089	114	0.179 152	7.315 897	1.466 564
40	RMSprop	0.000 082	570	0.471 317	3.003 355	1.469 744
46	RMSprop	0.000 059	153	0.295 671	7.276 354	1.472 631
16	Adam	0.000 133	225	0.521 078	5.03 229	1.477 234
6	Adam	0.000 097	225	0.467 451	5.42 483	1.478 946
35	RMSprop	0.000 179	66	0.540 199	7.791 814	1.485 383
2	RMSprop	0.000 052	225	0.487 136	7.756 889	1.485 913
0	RMSprop	0.000 075	225	0.505 513	6.664 503	1.488 749
15	Adam	0.000 035	225	0.436 437	5.701 631	1.490 682
37	RMSprop	0.000 151	73	0.560 933	6.776 026	1.495 198
28	RMSprop	0.000 035	265	0.364 326	4.117 896	1.497 929
19	Adam	0.000 144	225	0.582 637	6.224 196	1.499 658
21	Adam	0.000 261	225	0.422 763	6.114 556	1.502 561
20	Adam	0.000 025	225	0.401 408	6.951 192	1.506 756
7	Adam	0.000 069	225	0.496 068	7.374 401	1.51 577
13	Adam	0.000 031	225	0.452 886	5.889 216	1.530 536
49	RMSprop	0.000 265	195	0.52 036	6.28 119	1.544 203
18	Adam	0.000 016	225	0.43 794	5.56 616	1.549 521
38	RMSprop	0.000 434	605	0.494 833	6.954 664	1.549 902
29	Adam	0.000 039	400	0.360 762	7.630 682	1.560 381
45	RMSprop	0.000 329	332	0.413 058	4.932 252	1.580 241
1	RMSprop	0.000 285	225	0.555 782	7.176 087	1.583 419
26	Adam	0.000 047	987	0.32 565	4.25 768	1.593 861
33	Adam	0.000 032	155	0.120 995	5.757 826	1.606 285
36	RMSprop	0.00 018	142	0.59 511	7.965 855	1.611 265
27	RMSprop	0.00 003	391	0.234 176	1.175 896	1.626 354
34	RMSprop	0.000 013	789	0.38 773	4.12 605	1.627 522
31	Adam	0.000 076	900	0.167 727	5.418 842	1.630 291
10	Adam	0.000 013	225	0.593 005	6.110 873	1.638 141
41	RMSprop	0.000 013	241	0.270 598	5.188 218	1.647 733
25	Adam	0.00 029	784	0.230 935	2.162 501	1.686 709
8	RMSprop	0.00 037	225	0.478 398	5.746 003	1.688 195
14	Adam	0.000 184	225	0.533 729	6.477 521	1.69 213
42	RMSprop	0.000 112	435	0.101 017	6.930 886	1.71 168
30	RMSprop	0.000 706	285	0.407 261	6.188 819	1.725 915
24	Adam	0.000 514	225	0.400 875	5.872 189	1.773 344
5	RMSprop	0.000 757	225	0.415 991	5.231 926	1.793 177
44	RMSprop	0.000 935	226	0.460 966	6.326 114	1.799 258
17	Adam	0.00 037	225	0.553 661	6.577 556	1.904 586
3	Adam	0.000 923	225	0.498 035	5.16 874	2.041 645
9	Adam	0.000 786	225	0.57 071	5.326 067	2.277 016

Appendix B. Extended data

In figure 5, we show a randomly selected example from the 100 examples that are averaged in figure 4. In this section, we show all other examples. There is no inherent physical difference between the figures shown here and figure 4—they simply correspond to different randomly-generated masks. The figures are vectorised, so they can be zoomed in digitally and explored at full resolution. It is evident from the figures presented here that most of them follow the same trend as figure 5.

ORCID iDs

Johannes Fiedler ⓘ https://orcid.org/0000-0002-2179-0625
Eivind Kristen Osestad ⓘ https://orcid.org/0009-0002-3332-0732
Bodil Holst ⓘ https://orcid.org/0000-0001-6809-2579

References

[1] Lawson R A and Robinson A P 2016 *Materials and Processes for Next Generation Lithography* 1st edn (Amsterdam: Elsevier)
[2] Berggren K K, Bard A, Wilbur J L, Gillaspy J D, Helg A G, McClelland J J, Rolston S L, Phillips W D, Prentiss M and Whitesides G M 1995 Microlithography by using neutral metastable atoms and self-assembled monolayers *Science* **269** 1255–7
[3] Baldwin K 2005 Metastable helium: atom optics with nano-grenades *Contemp. Phys.* **46** 105–20
[4] Ueberholz B, Kuhr S, Frese D, Gomer V and Meschede D 2002 Cold collisions in a high-gradient magneto-optical trap *J. Phys. B: At. Mol. Opt. Phys.* **35** 4899–914
[5] Colli L 1954 Ultraviolet photons in the decay of metastable argon atoms *Phys. Rev.* **95** 892–4
[6] Gardner J R, Anciaux E M and Raizen M G 2017 Communication: neutral atom imaging using a pulsed electromagnetic lens *J. Chem. Phys.* **146** 081102
[7] Adams C S, Sigel M and Mlynek J 1994 Atom optics *Phys. Rep.* **240** 143–210
[8] Hinderthür H, Pautz A, Ruschewitz F, Sengstock K and Ertmer W 1998 Atom interferometry with polarizing beam splitters *Phys. Rev. A* **57** 4730–5
[9] Nesse T, Simonsen I and Holst B 2019 Nanometer-resolution mask lithography with matter waves: near-field binary holography *Phys. Rev. Appl.* **11** 024009
[10] Eder S D, Ravn A K, Samelin B, Bracco G, Salvador Palau A, Reisinger T, Knudsen E B, Lefmann K and Holst B 2017 Zero-order filter for diffractive focusing of de Broglie matter waves *Phys. Rev. A* **95** 023618
[11] Koch M, Rehbein S, Schmahl G, Reisinger T, Bracco G, Ernst W E and Holst B 2008 Imaging with neutral atoms-a new matter-wave microscope *J. Microsc.* **229** 1–5
[12] Doak R B, Grisenti R E, Rehbein S, Schmahl G, Toennies J P and Wöll C 1999 Towards realization of an atomic de Broglie microscope: helium atom focusing using fresnel zone plates *Phys. Rev. Lett.* **83** 4229–32
[13] Carnal O, Sigel M, Sleator T, Takuma H and Mlynek J 1991 Imaging and focusing of atoms by a fresnel zone plate *Phys. Rev. Lett.* **67** 3231–4
[14] Eder S D, Guo X, Kaltenbacher T, Greve M M, Kalläne M, Kipp L and Holst B 2015 Focusing of a neutral helium beam with a photon-sieve structure *Phys. Rev. A* **91** 043608

[15] Eder S D, Reisinger T, Greve M M, Bracco G and Holst B 2012 Focusing of a neutral helium beam below one micron *New J. Phys.* **14** 073014

[16] Patton F S, Deponte D P, Elliott G S and Kevan S D 2006 Speckle patterns with atomic and molecular de Broglie waves *Phys. Rev. Lett.* **97** 013202

[17] Barr M, Fahy A, Martens J, Jardine A P, Ward D J, Ellis J, Allison W and Dastoor P C 2016 Unlocking new contrast in a scanning helium microscope *Nat. Commun.* **7** 10189

[18] Reisinger T and Holst B 2008 Neutral atom and molecule focusing using a fresnel zone plate *J. Vac. Sci. Technol.* B **26** 2374–9

[19] Fiedler J, Broer W and Scheel S 2017 Reconstruction of Casimir–Polder interactions from matter-wave interference experiments *J. Phys. B: At. Mol. Opt. Phys.* **50** 155501

[20] Fujita J, Morinaga M, Kishimoto T, Yasuda M, Matsui S and Shimizu F 1996 Manipulation of an atomic beam by a computer-generated hologram *Nature* **380** 691–4

[21] Lohmann A W and Paris D P 1967 Binary fraunhofer holograms, generated by computer *Appl. Opt.* **6** 1739–48

[22] Onoe M and Kaneko M 1979 Computer generated pure binary hologram *Electron. Commun. Jpn.* **62** 118–26

[23] Nesse T, Banon J-P, Holst B and Simonsen I 2017 Optimal design of grid-based binary holograms for matter-wave lithography *Phys. Rev. Appl.* **8** 024011

[24] Fiedler J and Holst B 2022 An atom passing through a hole in a dielectric membrane: impact of dispersion forces on mask-based matter-wave lithography *J. Phys. B: At. Mol. Opt. Phys.* **55** 025401

[25] Scheel S and Buhmann S Y 2008 Macroscopic QED - concepts and applications *Acta Phys. Slovaca* **58** 675–809

[26] Fiedler J *et al* 2023 Perspectives on weak interactions in complex materials at different length scales *Phys. Chem. Chem. Phys.* **25** 2671–705

[27] Buhmann S Y, Scheel S, Ellingsen S Å, Hornberger K and Jacob A 2012 Casimir–Polder interaction of fullerene molecules with surfaces *Phys. Rev. A* **85** 042513

[28] Arndt M, Nairz O, Vos-Andreae J, Keller C, van der Zouw G and Zeilinger A 1999 Wave-particle duality of C60 molecules *Nature* **401** 680–2

[29] Brand C, Fiedler J, Juffmann T, Sclafani M, Knobloch C, Scheel S, Lilach Y, Cheshnovsky O and Arndt M 2015 A Green's function approach to modeling molecular diffraction in the limit of ultra-thin gratings *Ann. Phys., Lpz.* **527** 580–91

[30] Gack N, Reitz C, Hemmerich J L, Könne M, Bennett R, Fiedler J, Gleiter H, Buhmann S Y, Hahn H and Reisinger T 2020 Signature of short-range van der waals forces observed in poisson spot diffraction with indium atoms *Phys. Rev. Lett.* **125** 050401

[31] Zhou L *et al* 2015 Test of equivalence principle at 10^{-8} level by a dual-species double-diffraction raman atom interferometer *Phys. Rev. Lett.* **115** 013004

[32] Gustavson T L, Bouyer P and Kasevich M A 1997 Precision rotation measurements with an atom interferometer gyroscope *Phys. Rev. Lett.* **78** 2046–9

[33] Keith D W, Ekstrom C R, Turchette Q A and Pritchard D E 1991 An interferometer for atoms *Phys. Rev. Lett.* **66** 2693–6

[34] Clauser J F and Li S 1994 Talbot-vonlau atom interferometry with cold slow potassium *Phys. Rev. A* **49** R2213–6

[35] Grisenti R E, Schöllkopf W, Toennies J P, Hegerfeldt G C and Köhler T 1999 Determination of atom-surface van der waals potentials from transmission-grating diffraction intensities *Phys. Rev. Lett.* **83** 1755–8

[36] Garcion C, Fabre N, Bricha H, Perales F, Scheel S, Ducloy M and Dutier G 2021 Intermediate-range casimir-polder interaction probed by high-order slow atom diffraction *Phys. Rev. Lett.* **127** 170402

[37] Arndt M, Hackermüller L and Reiger E 2005 Interferometry with large molecules: exploration of coherence, decoherence and novel beam methods *Braz. J. Phys.* **35** 216–23

[38] Hemmerich J L, Bennett R, Reisinger T, Nimmrichter S, Fiedler J, Hahn H, Gleiter H and Buhmann S Y 2016 Impact of Casimir–Polder interaction on Poisson-spot diffraction at a dielectric sphere *Phys. Rev. A* **94** 1–11

[39] Aster R C, Borchers B and Thurber C H (eds) 2013 *Parameter Estimation and Inverse Problems* 2nd edns (Boston: Academic)

[40] Born M and Wolf E 1999 *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light* 7th edn (Cambridge: Cambridge University Press)

[41] Braat J J M, van Haver S, Janssen A J E M and Dirksen P 2008 *Assessment of Optical Systems by Means of Point-Spread Functions* (*Progress in Optics*) vol 51, ed E Wolf (Amsterdam: Elsevier) pp 349–468

[42] Goodfellow I, Bengio Y and Courville A 2016 *Deep Learning* (Cambridge: MIT press)

[43] Gonzalez A and Perez R 1996 A learning system of fuzzy control rules based on genetic algorithms *Genetic Algorithms and Soft Computing* (Heidelberg: Physica-Verlag) pp 202–25

[44] Hornik K, Stinchcombe M and White H 1989 Multilayer feedforward networks are universal approximators *Neural Netw.* **2** 359–66

[45] Chen X, Wei Z, Li M and Rocca P 2020 A review of deep learning approaches for inverse scattering problems (invited review) *Prog. Electromagn. Res.* **167** 67–81

[46] Ren Y, Zhang L, Wang W, Wang X, Lei Y, Xue Y, Sun X and Zhang W 2021 Genetic-algorithm-based deep neural networks for highly efficient photonic device design *Photon. Res.* **9** B247–52

[47] Nigam A K, Pollice R and Aspuru-Guzik A 2021 JANUS: parallel tempered genetic algorithm guided by deep neural networks for inverse molecular design (arXiv:2106.04011)

[48] Köker R 2013 A genetic algorithm approach to a neural-network-based inverse kinematics solution of robotic manipulators based on error minimization *Inf. Sci.* **222** 528–43

[49] Zhao M, Chen Z and Blaabjerg F 2009 Optimisation of electrical system for offshore wind farms via genetic algorithm *IET Renew. Power Gener.* **3** 205–16

[50] Wei Z and Chen X 2018 Deep-learning schemes for full-wave nonlinear inverse scattering problems *IEEE Trans. Geosci. Remote Sens.* **57** 1849–60

[51] Orhan A E and Pitkow X 2017 Skip connections eliminate singularities (arXiv:1701.09175)

[52] Jadon S 2020 A survey of loss functions for semantic segmentation *2020 IEEE Conf. on Computational Intelligence in Bioinformatics and Computational Biology* (*CIBCB*) (IEEE) pp 1–7

[53] Kingma D P and Ba J 2014 Adam: a method for stochastic optimization (arXiv:1412.6980)

[54] Hinton G, Srivastava N and Swersky K 2012 Neural networks for machine learning lecture 6a overview of mini-batch gradient descent *Cited on* **14** 2

[55] Eiben A E *et al* 2003 *Introduction to Evolutionary Computing* vol 53 (Berlin: Springer)

[56] Gad A F 2021 Pygad: an intuitive genetic algorithm python library (arXiv:2106.06158)

[57] Messiah A 2014 *Quantum Mechanics* (*Dover Books on Physics*) (Mineola, New York: Dover Publications)

[58] Landau L D and Lifshitz E M 1975 *The Classical Theory of Fields* vol 2, 4th edn (Oxford: Pergamon)

[59] Kammler D W 2008 *A First Course in Fourier Analysis* 2nd edn (Cambridge: Cambridge University Press)

[60] Palau Aà S, Eder S D, Andersen T, Ravn A K, Bracco G and Holst B 2018 Center-line intensity of a supersonic helium beam *Phys. Rev.* A **98** 063611