

Scientific intuition inspired by machine learning generated hypotheses

Pascal Friederich,^{1, 2, 3, 4} Mario Krenn,^{1, 2, 5} Isaac Tamblyn,^{6, 5} and Alán Aspuru-Guzik^{1, 2, 5, 7}

¹*Chemical Physics Theory Group, Department of Chemistry, University of Toronto, Canada.*

²*Department of Computer Science, University of Toronto, Canada.*

³*Institute of Theoretical Informatics, Karlsruhe Institute of Technology,
Am Fasanengarten 5, 76131 Karlsruhe, Germany.*

⁴*Institute of Nanotechnology, Karlsruhe Institute of Technology,
Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany.*

⁵*Vector Institute for Artificial Intelligence, Toronto, Canada.*

⁶*National Research Council of Canada, Ottawa, Canada.*

⁷*Canadian Institute for Advanced Research (CIFAR) Lebovic Fellow, Toronto, Canada*

(Dated: October 28, 2020)

Abstract Machine learning with application to questions in the physical sciences has become a widely used tool, successfully applied to classification, regression and optimization tasks in many areas. Research focus mostly lies in improving the accuracy of the machine learning models in numerical predictions, while scientific understanding is still almost exclusively generated by human researchers analysing numerical results and drawing conclusions. In this work, we shift the focus on the insights and the knowledge obtained by the machine learning models themselves. In particular, we study how it can be extracted and used to inspire human scientists to increase their intuitions and understanding of natural systems. We apply gradient boosting in decision trees to extract human interpretable insights from big data sets from chemistry and physics. In chemistry, we not only rediscover widely known rules of thumb but also find new interesting motifs that tell us how to control solubility and energy levels of organic molecules. At the same time, in quantum physics, we gain new understanding on experiments for quantum entanglement. The ability to go beyond numerics and to enter the realm of scientific insight and hypothesis generation opens the door to use machine learning to accelerate the discovery of conceptual understanding in some of the most challenging domains of science.

I. INTRODUCTION

Machine learning (ML) recently became a widely used tool with many applications in the physical sciences [1], ranging from chemistry (for example, prediction of quantum chemistry properties [2] or inverse materials design [3, 4]) to physics (identification of phases of matter [5] or astronomical object recognition [6]) and biology (for example, prediction of protein structures [7] or drug design [8, 9]). While ML models are successfully used to accelerate predictions or to recognize or generate patterns, it is rarely inquired how the machine finds solutions. Thus, the scientific insight obtained by the model is not directly transferred to human scientists. First attempts to use artificial intelligence in physical sciences aimed to directly answer scientific questions, *e.g.* determine the location of protein encodings in the genome [10]. Further attempts to employ machine learning models to obtain insight and help scientists to develop theories were focused on rediscovering solutions to already solved problems, *e.g.* to rediscover the coordinate transformation in astrophysical [11] and nonlinear dynamical systems [12], or to detect symmetries and conservation laws [13]. The methods used in these cases enforce information bottlenecks or interpretable transformations in the ML model that then can inspire scientific understanding for human researchers [14].

In this work, we propose to use machine learning and systematic data analysis to automate further the process of generation of human interpretable scientific hypotheses. We demonstrate the applicability of the approach using two questions in the natural sciences - a rediscovery task of chemistry knowledge (molecular energy levels and hydrophobicity) and the discovery of new intuitions in physics (quantum optics). We show that our approach “redisCOVERS” but also extends known chemical rules of thumb for solubility and energy levels of organic molecules with application in organic photovoltaics and organic light-emitting diodes and helps us to better understand the entanglement created in quantum optical experiments.

Our model represents its findings in a graph representation which is directly related to chemical or physical instances in the specific scientific domain. The results are statements regarding distinct subgraphs that can easily be comprehended and therefore, scientifically interpreted and understood by human experts. This is in stark contrast to conventional machine learning models where the internal representations are only indirectly connected with the real physical entities and thus hard to impossible to interpret.

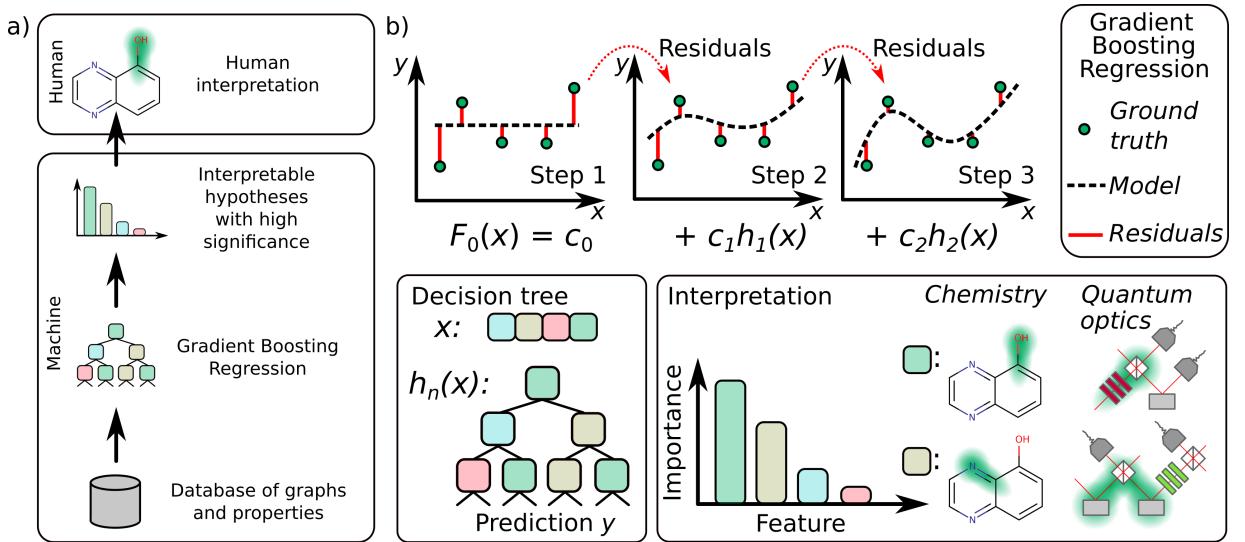


FIG. 1. Workflow for automated hypothesis generation. a) General workflow, starting with a database of graphs and respective properties, followed by training of a machine learning model that allows for the extraction of feature importances, *e.g.* Gradient Boosting Regression. Features with high importance are combined and analysed in a way that facilitates interpretation by researchers in order to stimulate scientific insight. b) Schematic illustration of the Gradient Boosting Regression method [15], where multiple simple decision tree models are trained sequentially, each new model correcting the residual errors of the previous models. The subgraph based input representation used in this work allows a direct interpretation of the feature importances that are computed based on a quantification of how meaningful features are for the accuracy of the machine learning model.

II. METHOD

Computer generated hypotheses. We suggest an automated workflow for ML-based generation of human interpretable scientific hypotheses as illustrated in Figure 1a. The workflow is based on a reference database of calculated (potentially also measured) data points with graph-based structure and with corresponding target properties. A binary feature vector describing presence/absence of automatically generated subgraphs [16] is used to train a tree ensemble method, *e.g.* Gradient Boosting [15] or Random Forrest Regression/Classification [17, 18], that allows for the quantification of feature importances. Based on the features with the highest importance, a list of hypotheses is generated. Each hypothesis has the human understandable form

“Feature i leads to an increase/decrease of target property of strength s ”

where i is the index of the corresponding feature (subgraph) in the input and strength s quantifies the degree of correlation between feature i and the target property. High feature importance does not necessarily correspond to a high direct correlation with the target feature. In many cases, multiple features have to be combined in order to become predictive, even if the single features individually do not help in the predicting the target property. Therefore, important features are combined

using logical operations (*and*, *xor*, ...) to automatically generate combined features which, especially in presence of higher-order correlations, can be directly interpreted by human researchers.

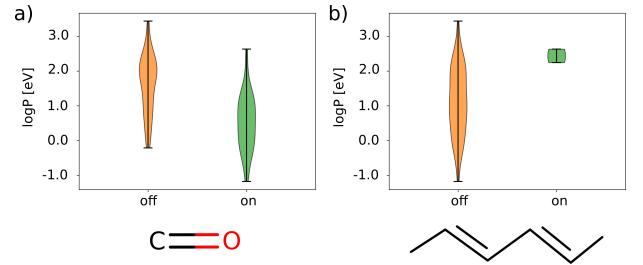


FIG. 2. Hypotheses about molecular solubility. a) Lower logP values (better solubility in water compared to octanol) can be achieved using the carbonyl groups, while b) conjugated carbon chains lead to higher logP values.

Input representation and experiments. In this work, we test this workflow on two experiments in chemistry and physics. The first experiment targets the automated generation of intuitive rules that determine molecular properties, whereas the second aims at hypothesis generation for entanglement properties of quantum optical experiments. In both cases, we can describe the data points as graphs (molecules and quantum optical experiments), where nodes are chemical elements or optical instruments while edges are chemical bonds or

photon paths travelling through the setup. This allows us to use fingerprinting techniques to generate input representations (bit-vectors), *e.g.* using the algorithm for circular extended-connectivity fingerprints [16]. This iterative algorithm generates a unique representation of each node, including its local environment. In each iteration, hashing functions are used to aggregate the information (predefined node and edge features) of the next nearest neighbors of each node, thus implicitly integrating information of one additional neighbor shell in each iteration. In the end, a hashing function is used to map all subgraphs found in the graphs to bit-vectors. Each entry in these bit-vectors encodes the presence or absence of a certain subgraph. A similar approach has been used in Lopez *et al.* [19] to determine molecular substructures in molecules for organic solar cells that lead to high power conversion efficiencies. Other models that link the presence of subgraphs (or more generally features) in the input data to properties can potentially be employed in our workflow (see e.g. Duvenaud *et al.* [20] where molecular fragments are identified that correlate with toxicity, the Grad-CAM method by Selvaraju *et al.* [21] for convolutional neural networks or the GNNExplainer by Ying *et al.* [22]). In contrast to this work, some of these approaches depend on the analysis of single samples and thus only indirectly allow to conclude about an entire data set. Furthermore, these approaches assign importance indicators to single nodes or edges of a graph, which are not necessarily binary numbers, which complicates the direct interpretation. Due to their general applicability to all graphs where node and edges can be represented by one or multiple categorical features, we focused on automatically generated circular fingerprints in this work.

III. RESULTS

To test the automated hypothesis generation workflow, we performed experiments in two scientific domains, molecular chemistry (Section III A) and quantum optical experiments (Section III B). We computed physical properties of these graphs and used the generated data sets and the workflow described in Figure 1 to automatically generate hypotheses that can be either compared to a collection of widely known chemical rules of thumb or that can help to better understand entanglement in quantum optical experiments for designing future experiments.

A. Chemical intuition for solubility, energy levels

In case of the chemistry experiment, we used two prototypical target properties - the water-octanol partition coefficient which describes the solubility of molecules

in water (polar) vs. octanol (non-polar) as well as the energy of the highest occupied molecular orbital. Both properties are of high relevance for the application of molecules as pharmaceuticals or in electronic devices, *e.g.* for organic solar cells, organic light-emitting diode (OLED) displays or organic flow batteries. We furthermore analysed existing application-specific data sets, namely a data set of thermally activated delayed fluorescent (TADF) molecules as emitter molecules for OLEDs [23], the Harvard Clean Energy project data set [24, 25] and a data set of non-fullerene acceptor molecules for organic solar cells [19]. Solubility and energy levels are relatively well understood and for both properties there exist several widely known rules of thumb, often described as chemical intuition, which describes how certain functional groups influence them. Our experiment aims to test whether the automated hypothesis generation method can “rediscover” those rules and potentially add new or refined rules. For frontier orbital gaps reported in the Harvard Clean Energy data set and the non-fullerene acceptor data set as well as for singlet-triplet energy splittings reported in the TADF data set, there exists less chemical intuition on how to influence and tune them.

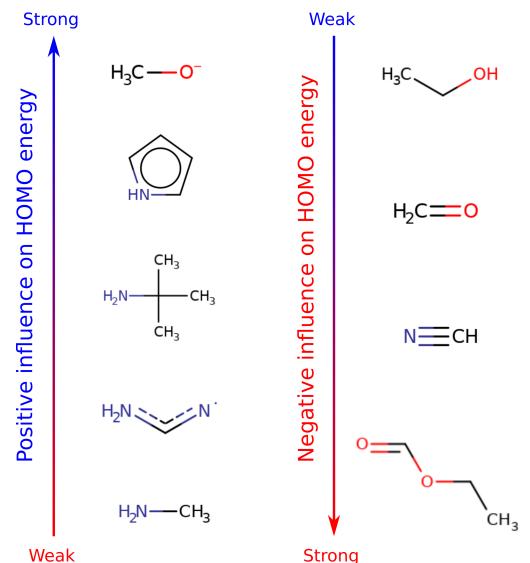


FIG. 3. Hypotheses about molecular energy levels. Molecular subgraphs with a positive (left) and negative (right) influence on the HOMO energy. The groups “discovered” by our automated workflow are widely known activating (resonance donating or electron donating) and deactivating groups, such as oxido/amino groups and nitrile groups.

Figure 2 shows two solubility related hypotheses that were generated using our workflow. Without prior knowledge, the algorithm predicts two widely known chemical groups/motifs for increasing solubility in polar solvents (carbonyl group in Figure 2a) and to increase solubility in non-polar solvents (conjugated carbon chain in Figure 2b). Figure 3 shows an overview of molecular

subgraphs that positively and negatively influence the HOMO energy of a molecule. To our surprise, five of the nine groups shown in the figure can directly be found in chemistry textbooks or Wikipedia when searching for electrophilic aromatic directing groups which can change the energy levels of molecules through the inductive effect and the mesomeric effect. Specifically, the oxido (O^-) group that shows the strongest positive influence on the HOMO is well known for a strong resonance donating and a strong inductive effect which both leads to an increase in HOMO energy. Furthermore, heterocycles that contain nitrogen, as well as amine (NH_2) groups are also known for lifting the HOMO level to higher energies. On the other hand, the nitrile group ($C\equiv N$) is one of the most widely known electron-withdrawing groups that lowers the HOMO energy of molecules due to its resonance withdrawing and inductively withdrawing nature.

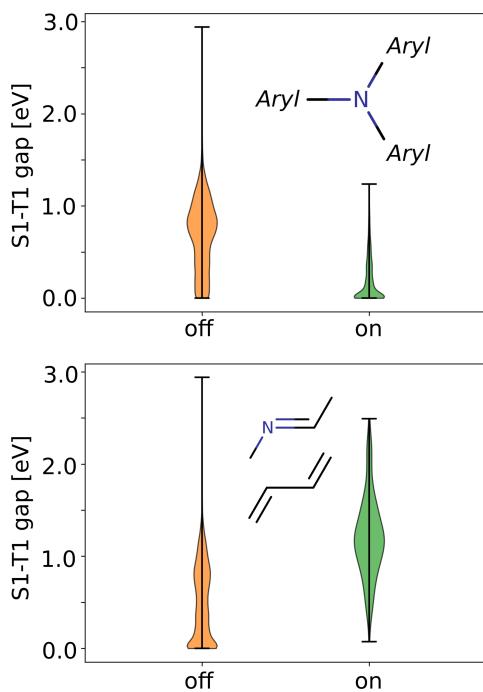


FIG. 4. Hypotheses singlet-triplet splittings in the TADF data set [23]. The data-driven algorithm finds the well known and widely exploited structure-property relation of triarylaminies and small single triplet gaps (<0.5 eV, upper panel). However, it finds an additional, less known motif of alternating single-double-bond bridges that are related to increased singlet triplet gaps (>0.5 eV, lower panel).

The patterns found to be relevant for small HOMO-LUMO gaps in the Harvard Clean Energy data set as well as in the non-fullerene acceptor data set are mostly related to extended aromatic systems and fused aromatic rings (see Figure S4a and Figure S1a). This finding is well-understood by chemists due to the widely known relation between the size of an aromatic system (*i.e.* the degree of delocalization of π -electrons) and the frontier orbital gap [26]. In the limit of infinite delocalization (*e.g.*

in graphene), the HOMO-LUMO gap closes completely. This relation was also exploited in the development of conductive polymers, which was awarded with the Nobel Prize in Chemistry in 2000 and which created the field of organic electronics [27].

However, we additionally found several interesting and surprising patterns both in the photovoltaic data sets (Figure S4b/c) and in the TADF dataset (Figure 4). In case of the Harvard Clean Energy data set, we find that aromatic heterocycles with sulfur (*e.g.* thiophene rings) as well as silicon heteroatoms (*e.g.* silole rings) significantly reduce the HOMO-LUMO gap. While the former are widely used in organic electronics to control energy levels and reduce HOMO-LUMO gaps, silole rings are more unusual.

In the non-fullerene acceptor data set (see Figure S4c) we found that thiophene rings connected by double bonds (*i.e.* forming a quinoid structure instead of aromatic systems) also significantly reduce the HOMO-LUMO gap, which is a known relation first described by Brédas [28]. However, such systems require a specific functionalization in the periphery of the molecule to enforce the quinoid structure of the two thiophene rings, which intrinsically is less stable and thus higher in energy than the aromatic structure.

In case of the TADF data set (see Figure 4), we found expected patterns such as triarylaminies that correlate with decreased singlet triplet gaps (S1-T1 gaps) as well as rather unexpected patterns (*e.g.* conjugated bridges) that are identified by our workflow as chemical groups that highly correlate with large singlet triplet gaps. Low singlet-triplet splittings in TADF molecules are typically achieved by decoupling electron donating and electron accepting parts of a molecule to reduce the exchange interaction between the frontier orbitals which would otherwise lower the triplet state compared to the singlet state and open an undesired singlet-triplet splitting. The decoupling of the fragments can be achieved by introducing twist angles close to 90° between the fragments. One way to accomplish this are triarylaminies bridges between the fragments. We expect that the conjugated bridges between fragments have precisely the opposite effect: They lead to a planar alignment of the adjacent fragments and thus an enhanced exchange interaction, reduced triplet energies and finally increased singlet-triplet splittings.

B. Physical intuitions for quantum experiments

As a second example, we use quantum optical experiments for producing high-dimensional, multipartite quantum entanglement [29, 30]. These experiments grow in interest as they allow the investigation of fundamental physical properties - such as local realism [31] - in laboratories. Furthermore, such quantum states are the key resources for large and complex quantum communication

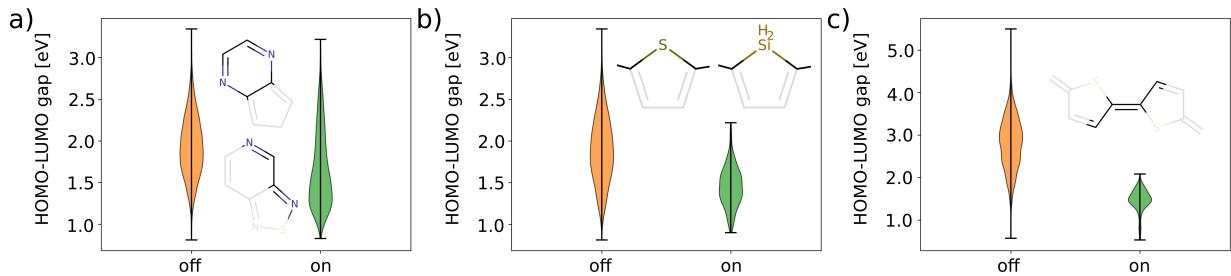


FIG. 5. Hypotheses about HOMO-LUMO gaps in the Harvard Clean Energy data set [24, 25] and a non-fullerene acceptor data set [19]. (a) The automated hypotheses generation protocol rediscovered the widely known relation between extended aromatic systems (containing *e.g.* nitrogen heteroatoms) and reduced HOMO-LUMO gaps. (b) Thiophene but also more uncommon silole rings are found to correlate with small HOMO-LUMO gaps. c) Thiophene rings bridged with double bonds (quinoid structures) are found to decrease the HOMO-LUMO gap in the non-fullerene acceptor data set. (Note the different scale in panel (c) compared to (a) and (b), due to differences in the data sets.)

networks [32, 33], which are on the edge of commercial availability. The experimental setups that we consider consist of standard optical components that are used in labs, such as nonlinear crystals for the creation of photon pairs, single-photon detectors, beam splitters, holograms or Dove prisms. Under approximations that are closely resembled in experiments, the final emergent quantum state can be reliably calculated [34].

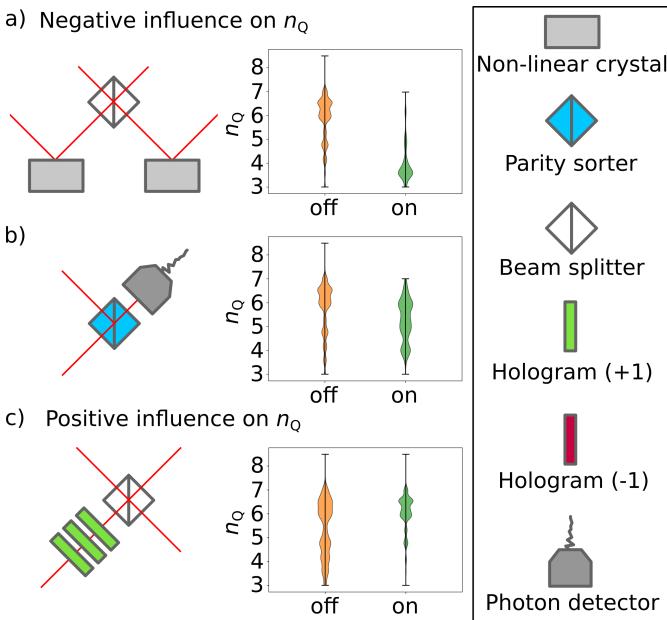


FIG. 6. Hypotheses about quantum optical experiments. Experimental substructures leading to a decrease in the overall size of the Hilbert space of involved qubits (n_Q) are shown in a) while substructures with positive influence are shown in b).

A key challenge lies in the design of experiments which creates certain desired quantum systems. The difficulty arises from counter-intuitive quantum phenomena, which raises the question of whether human intuition is the best

way to design new experiments. Several studies have therefore developed automated and machine-learning augmented approaches for the design of experiments [35–40]. The goal in our approach is to tackle this challenge in a completely different way, namely by improving the human scientist’s intuition about these systems.

Specifically, we are investigating optical setups with three-photon entanglement in high dimensions, using a fourth photon as a trigger. The experimental setups can be represented as graphs where vertices represent optical elements, and edges correspond to the photon paths connecting these elements. Analogously to chemical elements, the optical elements can have one to four connections. For example, a beam splitter has four input-output modes, while a detector has only one input. As a measure of entanglement, we use the overall size of the involved Hilbert space in terms of involved qubits, $n_Q = \log_2(d_1 d_2 d_3)$, where d_i stands for the rank of density matrix after tracing out photon i [41, 42].

We used the same fingerprint-based graph representation as in Section III A and trained a Gradient Boosting Regression model to predict n_Q . Using the algorithm outlined in Figure 1, we form a list of hypotheses of subgraphs features that influence n_Q most. This computer-generated list was analysed and interpreted by a human domain expert.

The two features which influence n_Q most negatively contradict the intuition in the field, see Fig. 6a/b and Fig. S2. Surprisingly, both of them represent subgraphs that are core elements of two experimental setups which have produced high-dimensional multipartite entanglement in the laboratory[43, 44]. Specifically, if the outputs of two nonlinear crystals (both crystals produce entangled photon pairs in the same 3-dimensional mode space) are connected directly via a beam splitter or interferometer, the entanglement of the resulting state is predicted to be comparably low. This can be interpreted

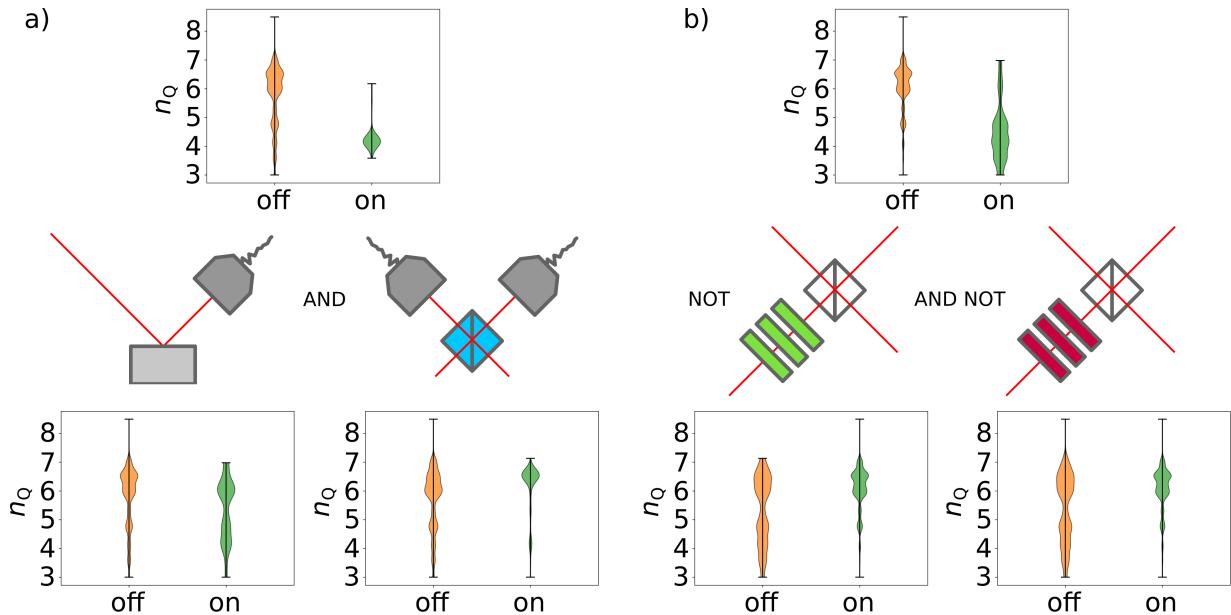


FIG. 7. Logically combined hypotheses about quantum optical experiments. Combining single-subgraph hypotheses with logical operations leads to intuitively interpretable relations, which is illustrated here with two examples. The upper panel shows the logically combined feature and its correlation with n_Q , while the lower panels show the correlations of the isolated subgraphs.

in the following way: The photons from the two different crystals need to combine at some point, otherwise, they remain bi-separable. However, if they combine directly after their generation, the equal mode spaces mix in such a way that it is difficult to increase their dimensionality subsequently. It is therefore explicitly enlightening that several of the features that positively influence n_Q correspond to elements which shift the entire mode space by plus or minus three before or after the beam splitters or nonlinear crystals. The insight for a human researcher now is to shift the mode space by three (as the local dimension is three), before combining photons from different nonlinear crystals to achieve a high n_Q . This leads to mode spaces of twice the original size and thereby increasing the probability for large overall entanglement dimensionalities.

A different feature which was used in the two experimentally demonstrations, but significantly negatively influences n_Q is the following: One output of a nonlinear crystal is directly connected to the detector. For human designers, this leads to the convenient fact that it simplifies the initial state (as double-emissions from one crystal can be ignored in this case). However, the entanglement of this photon with the other two photons can never be larger than three (as the local mode space is three). A similar, negatively influencing feature is a certain interferometer, which sorts the parity of the involved modes, directly connected to a detector. This acts as a filter, thus reducing the mode space of the incoming photon by half, thereby reducing the overall possible entanglement significantly.

Logically combined features: We can logically combine graph features, as described in II, and find the most significant macro-features for quantum experiments. In Fig. 7a, two small sub-experiments are combined with a logical *and*, *i.e.* the feature is the combination of both structures. Individually, the presence of the first feature has a negative influence on n_Q . The second feature, a parity sorter followed by two detectors, influences n_Q positively. Surprisingly, their combination has a significant negative influence on n_Q and can be seen as an almost sufficient condition for $n_Q \approx 4$. This behaviour can be interpreted using the Klyshko advanced wave-picture for quantum correlations in quantum optics [45]. The detector after the photon pair creation heralds a specific quantum state in the other photonic path. If those photons deterministically split at the parity sorter, the ability to mix with the photons from the other input ports (thus from the other crystal) vanish. From this insight, the human designer can learn that a heralded single-photon should be combined in a probabilistic way with the photons of the other crystal, using beam splitters instead of parity sorters.

A second macro-feature, Fig. 7b, combines two insights that we gained in Fig. 6. The macro-feature in Fig. 7b shows that the absence of either three positive or three negative mode shifters in front of a beam splitter has a very negative impact on the n_Q . Thereby, the algorithm has discovered that both increasing or decreasing helps to have very positive influence on the final entanglement, and thereby suggests that one can be agnostic about

the shift direction, and the importance lies in the actual increase of the local Hilbert space before the mixing. This features clearly shows how logical combinations can simplify the interpretation of scientific data.

IV. CONCLUSION AND OUTLOOK

We presented a data-driven machine learning workflow for automated generation and verification of hypotheses about observations in natural sciences. We presented examples from chemistry and physics, but our method is directly applicable to most applications, where structures can be represented as graphs, *e.g.* to DNA/RNA data in biology [46, 47], chemical reaction networks [48, 49] or graphs in social sciences. In chemistry, the workflow “redisCOVERS” widely known relations regarding solubility and electronic properties of molecules (often referred to as chemical intuition). In physics, the algorithm discovers rules to generate highly entangled three-photon states in quantum optical experiments. These rules are interpretable by human experts in retrospect, yet not known or postulated before, and even contradicting some of the field’s current understanding. Finding such rules will not only help human researchers to understand complex scientific relationships and thus design better experiments, but also reduce unavoidable and often undetectable bias generated by prior knowledge and anticipations.

Hypothesis testing. In addition to automated hypothesis generation, protocols for testing of the postulated hypotheses would be beneficial. In case of the chemistry experiment, a possible hypothesis testing protocol would

generate mutations of each molecule in the training set to test the hypotheses on molecules with similar representations, where (ideally) only the relevant feature is changed. In case of the quantum optical experiments, not all random mutations will lead to maximally entangled states between all photons, which is a requirement to compute the entanglement of the quantum state. We currently see two options for automated hypothesis verification both of which we are currently implementing. The first follows the same procedure of mutation and computation as in the chemistry experiment, with the caveat that only a small fraction of the mutations will lead to useful results, potentially making the procedure computationally costly. The second option is based on finding other experimental setups within the whole database that are as similar to the reference experiment as possible, with the exception of the feature that is currently analysed. This procedure is computationally costly as well but does not require new computations.

ACKNOWLEDGEMENTS

P.F. acknowledges funding the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 795206 (MolDesign). M.K. acknowledges support from the Austrian Science Fund (FWF) through the Erwin Schrödinger fellowship No. J4309. I.T. acknowledges NSERC and performed work at the NRC under the auspices of the AI4D and MCF Programs. A.A.-G. thanks Anders G. Frøseth for his generous support. A.A.-G. acknowledges the generous support of Natural Resources Canada and the Canada 150 Research Chairs program.

-
- [1] Giuseppe Carleo, Ignacio Cirac, Kyle Cranmer, Laurent Daudet, Maria Schuld, Naftali Tishby, Leslie Vogt-Maranto, and Lenka Zdeborová. Machine learning and the physical sciences. *Reviews of Modern Physics*, 91(4):045002, 2019.
 - [2] Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole von Lilienfeld. Big data meets quantum chemistry approximations: The δ -machine learning approach. *Journal of chemical theory and computation*, 11(5):2087–2096, 2015.
 - [3] Piotr S Gromski, Alon B Henson, Jarosław M Granda, and Leroy Cronin. How to explore chemical space using algorithms and automation. *Nature Reviews Chemistry*, 3(2):119–128, 2019.
 - [4] Benjamin Sanchez-Lengeling and Alán Aspuru-Guzik. Inverse molecular design using machine learning: Generative models for matter engineering. *Science*, 361(6400):360–365, 2018.
 - [5] Juan Carrasquilla and Roger G Melko. Machine learning phases of matter. *Nature Physics*, 13(5):431–434, 2017.
 - [6] Yashar D Hezaveh, Laurence Perreault Levasseur, and Philip J Marshall. Fast automated analysis of strong gravitational lenses with convolutional neural networks. *Nature*, 548(7669):555–557, 2017.
 - [7] Andrew W Senior, Richard Evans, John Jumper, James Kirkpatrick, Laurent Sifre, Tim Green, Chongli Qin, Augustin Žídek, Alexander WR Nelson, Alex Bridgland, et al. Improved protein structure prediction using potentials from deep learning. *Nature*, pages 1–5, 2020.
 - [8] Alex Zhavoronkov, Yan A Ivanenkov, Alex Aliper, Mark S Veselov, Vladimir A Aladinskiy, Anastasiya V Aladinskaya, Victor A Terentiev, Daniil A Polykovskiy, Maksim D Kuznetsov, Arip Asadulaev, et al. Deep learning enables rapid identification of potent ddr1 kinase inhibitors. *Nature biotechnology*, 37(9):1038–1040, 2019.
 - [9] Jonathan M Stokes, Kevin Yang, Kyle Swanson, Wengong Jin, Andres Cubillos-Ruiz, Nina M Donghia, Craig R MacNair, Shawn French, Lindsey A Carfrae, Zohar Bloom-Ackerman, et al. A deep learning approach to antibiotic discovery. *Cell*, 180(4):688–702, 2020.

- [10] Ross D King, Jem Rowland, Stephen G Oliver, Michael Young, Wayne Aubrey, Emma Byrne, Maria Liakata, Magdalena Markham, Pinar Pir, Larisa N Soldatova, et al. The automation of science. *Science*, 324(5923):85–89, 2009.
- [11] Raban Iten, Tony Metger, Henrik Wimling, Lídia Del Rio, and Renato Renner. Discovering physical concepts with neural networks. *Physical Review Letters*, 124(1):010508, 2020.
- [12] Bethany Lusch, J Nathan Kutz, and Steven L Brunton. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature communications*, 9(1):1–10, 2018.
- [13] Sebastian J. Wetzel, Roger G. Melko, Joseph Scott, Maysum Panju, and Vijay Ganesh. Discovering symmetry invariants and conserved quantities by interpreting siamese neural networks. *arXiv:2003.04299*, 2020.
- [14] Ribana Roscher, Bastian Bohn, Marco F Duarte, and Jochen Gericke. Explainable machine learning for scientific insights and discoveries. *IEEE Access*, 8:42200–42216, 2020.
- [15] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.
- [16] David Rogers and Mathew Hahn. Extended-connectivity fingerprints. *Journal of chemical information and modeling*, 50(5):742–754, 2010.
- [17] Tin Kam Ho. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1, pages 278–282. IEEE, 1995.
- [18] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [19] Steven A Lopez, Benjamin Sanchez-Lengeling, Julio de Goes Soares, and Alán Aspuru-Guzik. Design principles and top non-fullerene acceptor candidates for organic photovoltaics. *Joule*, 1(4):857–870, 2017.
- [20] David K Duvenaud, Dougal Maclaurin, Jorge Iparraguirre, Rafael Bombarell, Timothy Hirzel, Alán Aspuru-Guzik, and Ryan P Adams. Convolutional networks on graphs for learning molecular fingerprints. In *Advances in neural information processing systems*, pages 2224–2232, 2015.
- [21] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.
- [22] Zhitao Ying, Dylan Bourgeois, Jiaxuan You, Marinka Zitnik, and Jure Leskovec. Gnnexplainer: Generating explanations for graph neural networks. In *Advances in neural information processing systems*, pages 9244–9255, 2019.
- [23] Rafael Gómez-Bombarelli, Jorge Aguilera-Iparraguirre, Timothy D Hirzel, David Duvenaud, Dougal Maclaurin, Martin A Blood-Forsythe, Hyun Sik Chae, Markus Einzinger, Dong-Gwang Ha, Tony Wu, et al. Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nature materials*, 15(10):1120–1127, 2016.
- [24] Johannes Hachmann, Roberto Olivares-Amaya, Sule Atahan-Evrden, Carlos Amador-Bedolla, Roel S Sánchez-Carrera, Aryeh Gold-Parker, Leslie Vogt, Anna M Brockway, and Alán Aspuru-Guzik. The harvard clean energy project: large-scale computational screening and design of organic photovoltaics on the world community grid. *The Journal of Physical Chemistry Letters*, 2(17):2241–2251, 2011.
- [25] Steven A Lopez, Edward O Pyzer-Knapp, Gregor N Simm, Trevor Lutzow, Kewei Li, Laszlo R Seress, Johannes Hachmann, and Alán Aspuru-Guzik. The harvard organic photovoltaic dataset. *Scientific data*, 3(1):1–7, 2016.
- [26] Renana Gershoni-Poran, Anuja P Rahalkar, and Amnon Stanger. The predictive power of aromaticity: quantitative correlation between aromaticity and ionization potentials and homo-lumo gaps in oligomers of benzene, pyrrole, furan, and thiophene. *Physical Chemistry Chemical Physics*, 20(21):14808–14817, 2018.
- [27] Seth C Rasmussen. 2000 nobel prize in chemistry. In *Acetylene and Its Polymers*, pages 125–132. Springer, 2018.
- [28] Jean-Luc Bredas. Relationship between band gap and bond length alternation in organic conjugated polymers. *The Journal of chemical physics*, 82(8):3808–3811, 1985.
- [29] Nicolai Friis, Giuseppe Vitagliano, Mehul Malik, and Marcus Huber. Entanglement certification from theory to experiment. *Nature Reviews Physics*, 1(1):72–87, 2019.
- [30] Manuel Erhard, Mario Krenn, and Anton Zeilinger. Advances in high-dimensional quantum entanglement. *Nature Reviews Physics*, 2:365, 2020.
- [31] Jay Lawrence. Mermin inequalities for perfect correlations in many-qutrit systems. *Physical Review A*, 95(4):042123, 2017.
- [32] Matej Pivoluska, Marcus Huber, and Mehul Malik. Layered quantum key distribution. *Physical Review A*, 97(3):032312, 2018.
- [33] Xiao-Min Hu, Wen-Bo Xing, Chao Zhang, Bi-Heng Liu, Matej Pivoluska, Marcus Huber, Yun-Feng Huang, Chuan-Feng Li, and Guang-Can Guo. Experimental creation of multi-photon high-dimensional layered quantum states. *arXiv:2001.06253*, 2020.
- [34] Jian-Wei Pan, Zeng-Bing Chen, Chao-Yang Lu, Harald Weinfurter, Anton Zeilinger, and Marek Žukowski. Multiphoton entanglement and interferometry. *Reviews of Modern Physics*, 84(2):777, 2012.
- [35] Mario Krenn, Mehul Malik, Robert Fickler, Radek Lapkiewicz, and Anton Zeilinger. Automated search for new quantum experiments. *Physical Review Letters*, 116(9):090405, 2016.
- [36] PA Knott. A search algorithm for quantum state engineering and metrology. *New Journal of Physics*, 18(7):073033, 2016.
- [37] Julius Wallnöfer, Alexey A Melnikov, Wolfgang Dür, and Hans J Briegel. Machine learning for long-distance quantum communication. *PRX Quantum*, 1(1):010301, 2020.
- [38] Xiang Zhan, Kunkun Wang, Lei Xiao, Zhihao Bian, Yongsheng Zhang, Barry C Sanders, Chengjie Zhang, and Peng Xue. Experimental quantum cloning in a pseudo-unitary system. *Physical Review A*, 101(1):010302, 2020.
- [39] Mario Krenn, Manuel Erhard, and Anton Zeilinger. Computer-inspired quantum experiments. *Nature Reviews Physics*, 2:649–661, 2020.
- [40] Mario Krenn, Jakob Kottmann, Nora Tischler, and Alán Aspuru-Guzik. Conceptual understanding through efficient inverse-design of quantum optical experiments. *arXiv:2005.06443*, 2020.

- [41] Marcus Huber and Julio I de Vicente. Structure of multidimensional entanglement in multipartite systems. *Physical Review Letters*, 110(3):030501, 2013.
- [42] Marcus Huber, Martí Perarnau-Llobet, and Julio I de Vicente. Entropy vector formalism and the structure of multidimensional entanglement in multipartite systems. *Physical Review A*, 88(4):042328, 2013.
- [43] Mehul Malik, Manuel Erhard, Marcus Huber, Mario Krenn, Robert Fickler, and Anton Zeilinger. Multi-photon entanglement in high dimensions. *Nature Photonics*, 10(4):248, 2016.
- [44] Manuel Erhard, Mehul Malik, Mario Krenn, and Anton Zeilinger. Experimental greenberger–horne–zeilinger entanglement beyond qubits. *Nature Photonics*, 12(12):759–764, 2018.
- [45] DN Klyshko. A simple method of preparing pure states of an optical field, of implementing the einstein–podolsky–rosen experiment, and of demonstrating the complementarity principle. *Soviet Physics Uspekhi*, 31(1):74, 1988.
- [46] Thomas Gerling, Klaus F Wagenbauer, Andrea M Neuner, and Hendrik Dietz. Dynamic dna devices and assemblies formed by shape-complementary, non-base pairing 3d components. *Science*, 347(6229):1446–1452, 2015.
- [47] Florian Praetorius, Benjamin Kick, Karl L Behler, Maximilian N Honemann, Dirk Weuster-Botz, and Hendrik Dietz. Biotechnological mass production of dna origami. *Nature*, 552(7683):84–87, 2017.
- [48] Oleg N Temkin, Andrew V Zeigarnik, and DG Bonchev. *Chemical reaction networks: a graph-theoretical approach*. CRC Press, 1996.
- [49] Dmitrij Rappoport, Cooper J Galvin, Dmitry Yu Zubarev, and Alán Aspuru-Guzik. Complex chemical reaction networks from heuristics-aided quantum chemistry. *Journal of chemical theory and computation*, 10(3):897–907, 2014.

Supplementary Information:
Scientific intuition inspired by machine learning generated hypotheses

S1. ADDITIONAL HYPOTHESIS ABOUT CHEMISTRY DATA SETS

Figure S1 shows additional features that influence the HOMO-LUMO gap of molecules in multiple data sets. Our workflow finds patterns that are commonly associated with a positive or negative influence on HOMO-LUMO gaps, but also patterns and groups such as silole rings.

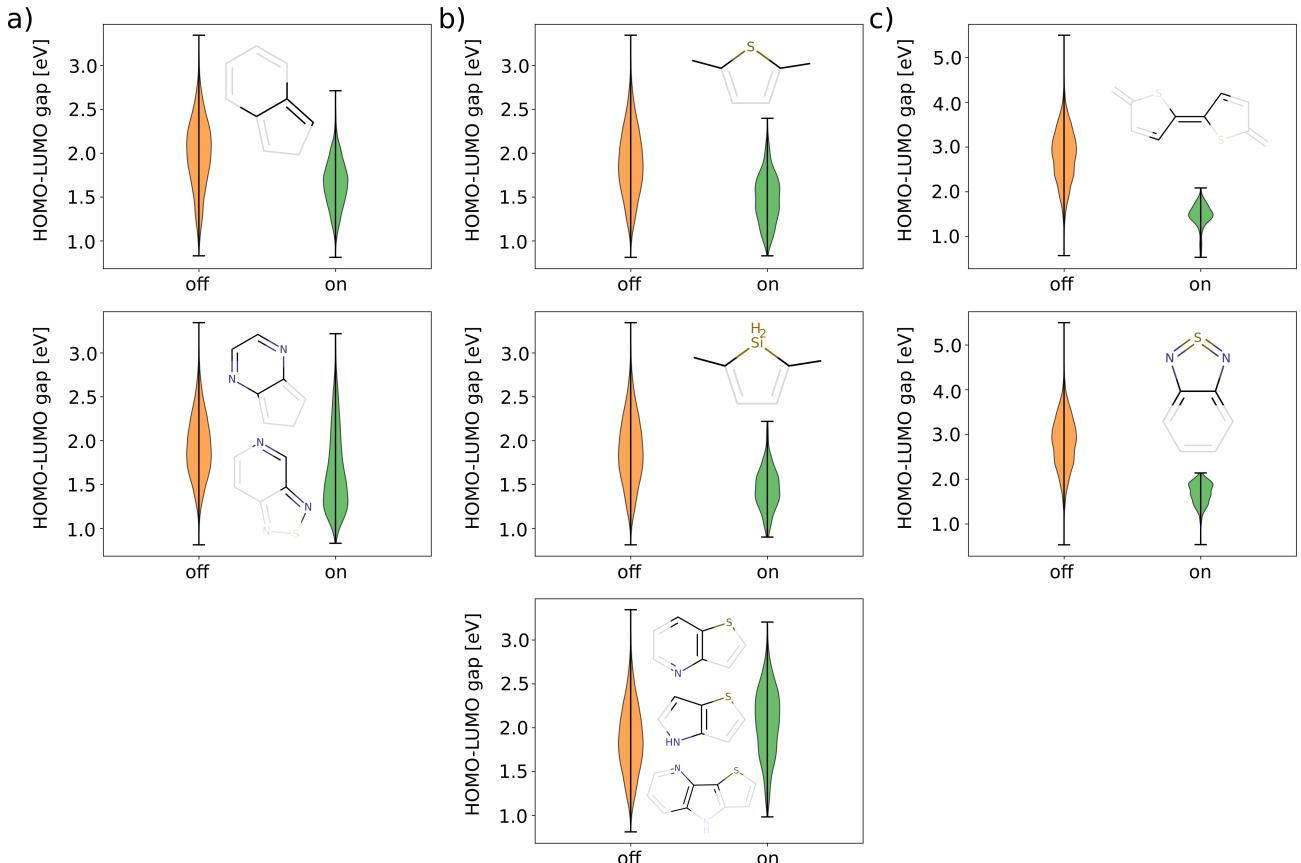


FIG. S1. HOMO-LUMO gaps in the Harvard Clean Energy data set [24, 25] and a non-fullerene acceptor data set.[19] (a) Agreeing with widely known rules of thumb, extended aromatic systems containing nitrogen heteroatoms are associated with reduced HOMO-LUMO gaps. (b) Thiophene but also more uncommon silole rings correlate with small HOMO-LUMO gaps. c) Thiophene rings bridged with double bonds (*i.e.* quinoid instead of aromatic systems) are found to decrease the HOMO-LUMO gap in the non-fullerene acceptor data set, a phenomenon that is studied and described in literature [28]. (Note the different scale in panel (c) compared to (a) and (b), due to differences in the data sets.)

Figure S2 shows additional features that influence the singlet-triplet gap of TADF molecules. Figure S2a shows groups with a negative influence on the singlet-triplet gap and Figure S2b shows groups with a positive influence on the singlet-triplet gap.

Figure S3 shows an example of a easily interpretable, logically combined feature that influences the singlet-triplet gap of molecules.

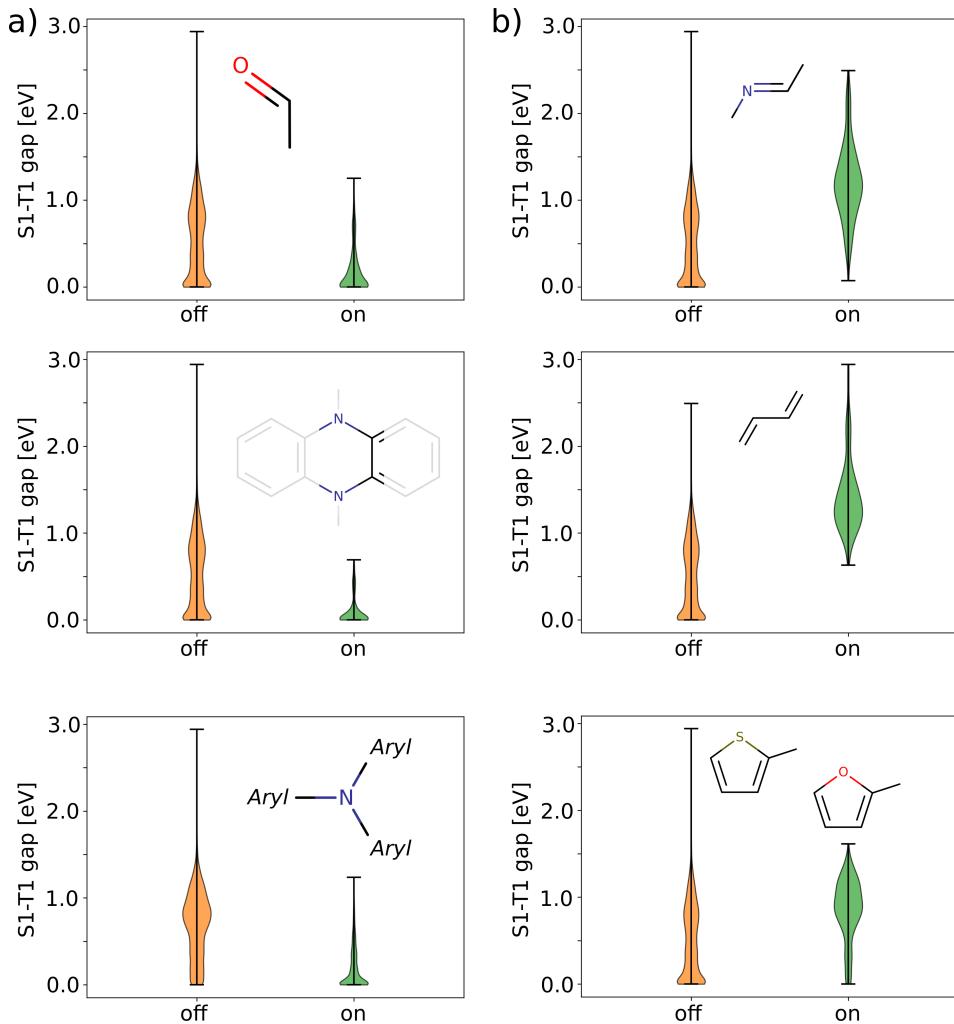


FIG. S2. **TADF** (a) Groups with a negative influence on the singlet-triplet gap, (b) Groups with a positive influence on the singlet-triplet gap.

S2. ADDITIONAL HYPOTHESIS ABOUT QUANTUM OPTICAL EXPERIMENTS

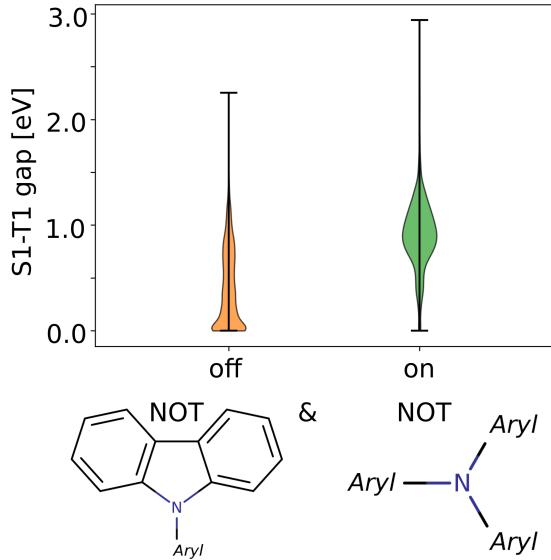
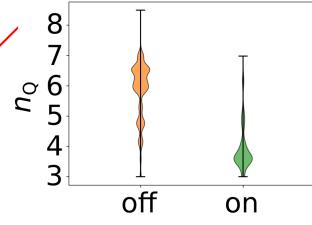
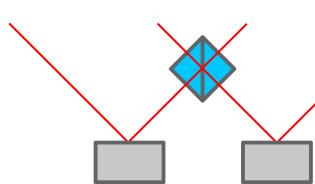


FIG. S3. **TADF** The logical combination of carbazole and triarylamine groups indicate that the algorithm identified those two groups as required for small singlet-triplet gaps, because the absence of both groups leads to significantly increased singlet-triplet gaps.

a) Negative influence on n_Q



b) Positive influence on n_Q

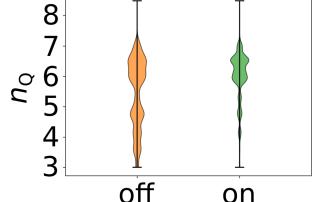
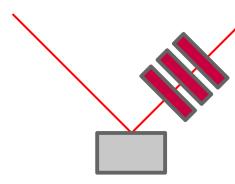
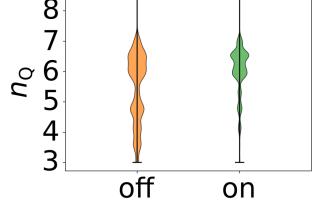
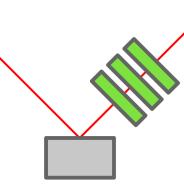
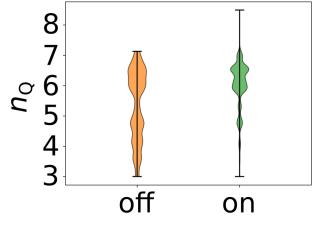
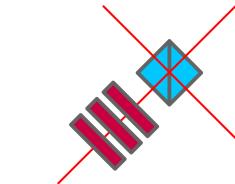
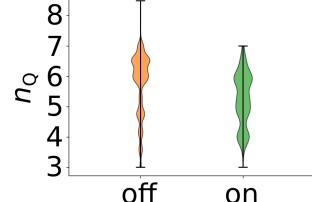
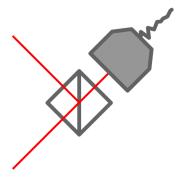
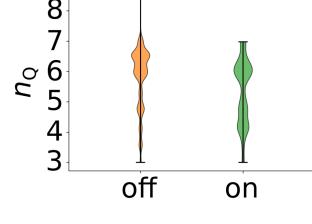
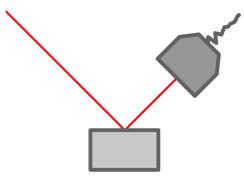
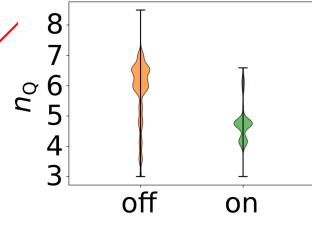
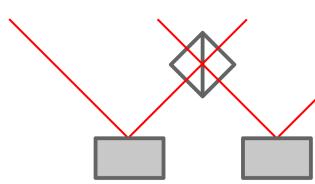
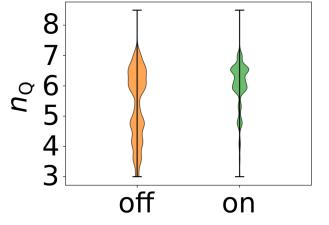
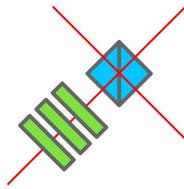


FIG. S4. **Entanglement in quantum optical experiments** Subgraphs of the experimental setups that a) decrease and (b) increase the overall size of the involved Hilbert space n_Q