Finding Molecules with Large Hyperpolarizabilities

Dominic Mashak, Steven A. Alexander^{*a*,*}

^aDepartment of Physics, Southwestern University, Georgetown, TX 78626, USA

alexands@southwestern.edu

(Received November 21, 2024)

Abstract

We describe a simulated annealing algorithm that can find molecules with large hyperpolarizabilities. This program represents each molecule as a SMILES string and modifies this string using seven mutation operators. We use the semi-empirical quantum chemistry program MOPAC to calculate the average hyperpolarizability. After a few iterations our algorithm significantly increased the value of this property.

1 Introduction

Many research groups are actively developing devices that are based on photons rather than electrons. This is because optical communication, optical computing, optical data storage and optical switching devices [6,9] should be much faster than their traditional counterparts. These optical devices depend on nonlinear optical (NLO) materials that dramatically modify light's frequency, phase and/or polarization. Organic molecules

This work is licensed under a Creative Commons "Attribution 4.0 International" license.



^{*}Corresponding author.

with large hyperpolarizabilities have garnered considerable attention because they have excellent frequency conversion and fast optical response time [2, 5, 19]. Experimental and theoretical groups can also tailor the properties of these molecules to particular applications.

In a recent paper, we showed that simulated annealing could find cyclic and acyclic molecules with specific boiling points and melting points [1]. This algorithm represented each molecule as a SMILES (simulated molecular input line entry specification) string [24] and assigned a fitness function value to it that quantified how close its properties were to some ideal values. This program used seven mutation operators to modify each SMILES string. Most modified molecules had a worse fitness function than the parent molecule, but a few were better. These better molecules then became the input to the next iteration. In this paper, we have modified the simulated annealing algorithm to find molecules with large average hyperpolarizabilities. Many quantum-chemical techniques can calculate this property, but here we use MOPAC, an open-source semi-empirical program [21] based on Dewar and Thiel's NDDO approximation [7]. Its linear scaling can quickly evaluate systems of up to 15000 atoms. In Section 2 we describe how we combined MOPAC with our simulated annealing program. Next, we apply our algorithm to ten molecules that a number of research groups have studied as promising NLO materials [10, 15, 20]. The results of these calculations are described in Section 3. Section 4 explores a variant of our simulated annealing algorithm that breaks a SMILES string into a fixed part that doesn't evolve and a second part that does. Because it superficially resembles the process used to construct integrated chips, we call this modified algorithm "masking".

2 An algorithm to find molecules with large β values

Simulated annealing is a numerical optimization method that can find the global minimum of complicated multidimensional functions [8,23]. Starting from some initial point, $x_{initial}$, and its value at that point, $f(x_{initial})$,

this method generates a new point in the multidimensional space, x_{new} , and calculates its value at that point, $f(x_{new})$. If $f(x_{initial}) > f(x_{new})$, this step is accepted and x_{new} becomes the starting point for the next step. If $f(x_{initial}) < f(x_{new})$, an acceptance function determines whether x_{new} is accepted or rejected. Simulated annealing tries to avoid getting stuck in a local minimum by occasionally accepting steps that yield worse solutions. In the acceptance function used by Metropolis et al. [17]

$$A = \min(R, exp([f(x_{new}) - f(x_{initial})]/T))$$
(1)

T is a parameter known as the temperature and R is a random number between 0.0 and 1.0. If the value of R is greater than the exponential, the new point becomes the initial point in the next step even though it has a worse fitness function. If R is less than the exponential, the previous point is retained to the next step.

In this paper we choose our fitness function, f(x), to be the molecule's average hyperpolarizability

$$\beta = \sqrt{\beta_x^2 + \beta_y^2 + \beta_z^2} \tag{2}$$

where

$$\beta_i = \frac{1}{3} \sum_j \beta_{ijj} + \beta_{jij} + \beta_{jji} \tag{3}$$

Here β_{ijk} are the components of the hyperpolarizability with i,j,k = (x,y,z) as computed by MOPAC [12]. In order to link MOPAC with our simulated annealing algorithm, several programs have to work together:

Step 1. The first part of our simulated annealing program reads a SMILES string and its fitness function value. This program then produces a new SMILES string from each of seven mutation operators:

* Pick a random bond and change it into a different type of bond (e.g. C=C-N-O \rightarrow C-C-N-O)

* Add a random atom with a random bond to a random location in the SMILES string (e.g. C=C-N-O \rightarrow C=C-O-N-O)

* Add a random atom with a random bond as a new branch to a random

location in the SMILES string (e.g. $C=C-N-O \rightarrow C=C-N(-O)-O$)

* Delete a random atom and its connecting bond (e.g.

 $C{=}C{-}N{-}O \rightarrow C{=}C{-}O)$

* Pick a random atom in the SMILES string and change it into a different type of atom (e.g. C=C-N-O \rightarrow C=C-N-C)

* Pick two random atoms in the SMILES string and connect them with a ring (e.g. C=C-N-O \rightarrow C1=C-N-O1)

* Delete a ring (e.g. C1=C-N-O1 \rightarrow C=C-N-C)

Step 2. The open-source program OpenBabel [18] converts each new SMILES string into an initial set of 3D coordinates for each atom, adds all hydrogen atoms and then creates a MOPAC input file.

Step 3. The open-source program MOPAC first optimizes the coordinates of each atom (GNORM=0.1) using the general-purpose force field PM6 [22]. At this minimum energy position, the hyperpolarizability is calculated without a solvent and with no frequency dependence (PO-LAR(E=0)).

Step 4. A short Python program reads the MOPAC output file and creates a file containing the extracted hyperpolarizability value.

Step 5. The second part of our simulated annealing program reads the fitness function values from all the new SMILES strings. The string with the best fitness function is then used in Eqn. 1 to determine whether the original string or this new string will become the starting point for the next step.

In Step 1, we immediately reject any molecule that violates the basic rules of molecular bonding (such as an oxygen atom with more than two single bonds). To create molecules that look relatively normal, we also reject any system with triple bonds and ringed systems that do not have 5 or 6 atoms. These requirements are somewhat arbitrary and they will exclude numerous valid molecules. We include them because they help avoid many systems that are obviously unphysical and this provides a relatively safe starting point for the calculations in the next section. Other restrictions can easily be added if needed.

3 Generating molecules with large β values

Several theoretical and experimental studies have identified numerous organic molecules that could form the basis of NLO materials. We have chosen ten of these molecules as the starting point for our simulated annealing algorithm. For testing purposes, we considered molecules that only contain carbon, oxygen, nitrogen and hydrogen atoms. We also restricted these molecules to those with single and double bonds and rings with 5 or 6 atoms. The first three entries in Table 1 are taken from a study by Marder et al. [15], the next four from a paper by Singer et al. [20] and the last three from a study by Kanis et al. [10].

Before performing a simulated annealing calculation on each initial molecule in Table 1, we must choose the temperature value in our acceptance function, Eqn. 1. If this temperature is too low, the optimization will quickly become trapped in a local minimum. If this temperature is too high, the optimization will not converge because the molecule will move randomly through the solution space. To find a viable working temperature, we performed a 100-step simulated annealing calculation on initial molecule #1 with three different temperatures, T=10, 20 and 40. As Figure 1 shows, T=20 gives the fastest convergence, so we used this value for the rest of our calculations.



Figure 1. The average hyperpolarizability of initial molecule #1 for 100 steps at three temperatures (T=10, 20 and 40).

Number	Initial/Final Molecule	β
1	C-N(-C)-C1=C-C=C(-C=C1)-N=C2-C=C-C(=O)-C=C2	3389.38
	C=C-N(-C)-C=C(-C)-C=C(-O(-N))-N=C1-N=C-C(=O)-C=N1	38635.99
2	C1=C-C(-C=C-C2-C=C-C(-C(-O)-O)-C=C2)=C-C=C1-N(-C)-C	1107.95
	C1-O-C(-C=C2-N=C(-C(-O)-O)-C=N2)=C=C=C1-N(-C=C)-C	18408.91
3	C1=C-C(-C=C-C=C-C=O)=C-C=C1-N(-C)-C	4137.17
	C-N-O-C(-C=C=C=C-N=O)=C=C-N(-C)-C(=C)-C(=O)	18811.46
4	C1=C-C(-N(-O)-O)=C-C=C1-N	470.46
	C=N-C(-N(-O)-O(-N=N1))=C1-C(-O)=C-N-N	5786.35
5	C1=C-C(-N(-N))=C-C=C1-N(-O)-O	693.03
	C1-N-C(-N(-C)(-N))=C(-O)-N=C1-C(=O)-N(-O(-C(=O)))	3009.90
6	C1=C-C(-C=C(-C)-C)=C-C=C1-N(-C)-C	567.03
	C(=N-O)=C-C(-N=C=C=N-O)=N-C=C-N(-C(=C))-C=C-N(-C)	10967.38
7	C-N(-C)-C1=C-C=C(-C=C1)-N=N-C2=C-C=C(-C=C2)-N(-O)-O	3656.56
	C-N(-C-C)-C=C-C=C(-C-C=O)-N=N-C=C=C=C(-N=O)-N(-O)	20300.14
8	N-C1=C-C=C(-C=C1)-C=C-C2=C-C=C(-C=C2)-N(-O)-O	1479.04
	N-C=C=C=C(-N)-C=C-C1=N-N=C(-N=N1)-N=C=C(-C=O)-N=O	20104.29
9	C1=C-C(-C=C-N(-O)-O)=C-C=C1-N(-C)-C	1423.70
	C1-C(-C=C(-O(-N(=C=O)))-O(-O))=C-C=C1-N(-C=C-C)-C	9275.78
10	C1=C-C(-C=C-C=C-N(-O)-O)=C-C=C1-N(-C)-C	2300.20
	C1-C-C(-C=C-N=C-N(-O)-O-N(=N))=C=C=C1-N(-N(=C))-C	12833.94

Table 1. Simulated annealing calculation using 100 steps and T=20

Most simulated annealing programs start at some high initial temperature. This normally allows a sequence of steps to effectively sample the whole parameter space since most solutions are accepted. After a certain number of steps, the temperature is repeatedly reduced. This should allow the optimization to avoid getting stuck in local minima and to eventually settle in the global minimum [23]. In this paper we are interested in improving the fitness function of each initial molecule, not in finding the molecule with the largest possible average hyperpolarizability. For this reason, we have simply performed a short 100-step calculation on each of the molecules in Table 1. On average, this process increased the average hyperpolarizability by a factor of 10. The smallest change (initial molecule #5) was 4.3 times and the largest (initial molecule #6) was 19.3 times. The structural shape of each final molecule is shown in Figure 2.



Figure 2. The structural shape of each final molecule in Table 1.

4 Generating molecules using masking

Because our simulated annealing algorithm modifies the entire SMILES string, the molecule at the end of this calculation looks very different than the molecule at the beginning. Many experimental and theoretical studies have taken a different approach - they pick a molecular framework and then modify parts of this structure to increase its hyperpolarizability. We can mimic this technique by breaking a SMILES string into a fixed part that doesn't evolve and a second part that does. Because it superficially resembles the process used to construct integrated chips, we call this modified algorithm "masking".



Figure 3. A generic representation of initial molecules #1, #2, #3, #6, #7, #9 and #10.

Table 2. Masking calculation using 100 steps and three temperatures (T=10, 50 and 100).

Т	Initial/Final Molecule	β
10	C-N(-C)-C1=C-C=C(-C=C1)-C=C(-C)-C	567.03
	C-N(-C)-C1=C-C=C(-C=C1)-N=C=C=O	2936.61
50	C-N(-C)-C1=C-C=C(-C=C1)-C=C(-C)-C	567.03
	C-N(-C)-C1=C-C=C(-C=C1)-N=C(-O(-O-O-O-O))-C(=O)-C(=O)-N=C=C-N=N	8912.85
100	C-N(-C)-C1=C-C=C(-C=C1)-C=C(-C)-C	567.03
	C-N(-C)-C1=C-C=C(-C=C1)-N=N-C(=O)-O(-O-N=N(-O-C(-O-O)=O))	6572.40

Most of the initial molecules in Table 1 contain an N,N-dimethylaniline structure. In Figure 3 we represent these molecules as a combination of this structure and a generic group R. If we pick initial molecule #6 to test our masking algorithm, the N,N-dimethylaniline structure becomes the fixed part (with a SMILES string of C-N(-C)-C1=C-C=C(-C=C1)-) and everything else becomes the flexible part (with a SMILES structure of C=C(-C)-C). Masking then modifies this flexible part with the seven mutation operators described in Section 2. Before OpenBabel and MOPAC evaluate these new SMILES strings, however, we reattach the fixed part to form a complete molecule.

In Table 2 we present the results of our masking algorithm using 100 steps and three different temperatures, T=10, 50 and 100. These calculations show that T=50 gives the fastest convergence. This temperature is higher than in the previous section because the fixed part of the molecule creates larger fluctuations in the fitness function. Despite this complication, our algorithm increased the average hyperpolarizability by 15.7 times. This amount is slightly smaller than the 19.3 times improvement

we obtained in the previous section using unconstrained optimization.

5 Conclusions

This paper shows that a simple simulated annealing algorithm can find molecules with large average hyperpolarizabilities. Our goal is not to find molecules with the largest possible hyperpolarizability but to show that substantial improvement is possible from some initial starting point. In these calculations, we took our initial molecules from three earlier studies that optimized the hyperpolarizability of several types of organic molecules [10, 15, 20]. After a relatively short calculation with only 100 steps, our algorithm increased the fitness function of each initial molecule between 4 and 19 times. We used the open-source program MOPAC because this open-source program can evaluate the hyperpolarizabilities of organic molecules quickly and relatively accurately [3, 14]. In a future paper, we plan to extend our algorithm to molecules that contain a more diverse set of atoms and to use more accurate Ab Initio methods such as Density Functional Theory [11, 13].

We also examined a variant of our simulated annealing algorithm that we call "masking". This modification evolves only part of a molecule; an ability that could be valuable in situations where there is a compelling theoretical or experimental need to have a common molecular core. As with our original algorithm, masking produced impressive gains in the fitness function.

Several studies (such as [4,16,19]) have described the difficulty of translating the properties of a single molecule into a useful macroscopic NLO material. The current consensus is that molecular hyperpolarizability is important, but the magnitude of the dipole moment, high thermal stability, and optical transparency at the operating wavelength are also essential properties (see, for example, [6]). For this reason, the calculations presented in this paper are a necessary first step toward identifying useful organic molecules that can eventually form NLO materials. In a forthcoming paper, we will include these other properties in our calculations using a modified fitness function. **Acknowledgment:** We profusely thank Mr. Sean Smith for his help during the early part of this investigation.

References

- S. A. Alexander, Generating molecules with specific boiling points and melting points, MATCH Commun. Math. Comput. Chem. 94 (2025) 77–93.
- [2] J. L. Bredas, Molecular geometry and nonlinear optics, Science 263 (1994) 487–488.
- [3] C. Cardoso, P. E. Abreu, B. F. Milne, F. Nogueira, Computational study of molecules with high intrinsic hyperpolarizabilities, *J. Phys. Chem. A* **114** (2010) 10676–10683.
- [4] L. R. Dalton, Nonlinear optical polymeric materials: from chromophore design to commercial applications, in: K. S. Lee (Ed.), *Polymers for Photonics Applications I*, Springer-Verlag, Berlin Heidelberg, 2002, pp. 1–86.
- [5] L. R. Dalton, Organic electro-optic materials, Pure Appl. Chem. 76 (2004) 1421–1433.
- [6] L. R. Dalton, P. A. Sullivan, D. H. Bale, Electric field poled organic electro-optic materials: state of the art and future prospects, *Chem. Rev.* **110** (2010) 25–55.
- [7] M. J. S. Dewar, W. A. Thiel, Semiempirical model for the two-center repulsion integrals in the NDDO approximation, *Theoret. Chim. Acta* 46 (1977) 89–104.
- [8] K. A. Dowsland, Simulated annealing, in: C. R. Reeves (Ed.), Modern Heuristic Techniques for Combinatorial Problems, Wiley, New York, 1993, pp. 20–69.
- [9] J. M. Hales, S. Barlow, H. Kim, S. Mukhopadhyay, J. L. Bredas, J. W. Perry, S. R. Marder, Design of organic chromophores for all-optical signal processing applications, *Chem. Mat.* 26 (2014) 549–560.
- [10] D. R. Kanis, M. A. Ratner, T. J. Marks, Design and construction of molecular assemblies with large second-order optical nonlinearities, *Chem. Rev.* 94 (1994) 195–242.

- [11] R. Kumar, S. K. Yadav, R. Seth, A. Singh, Designing of gigantic firstorder hyperpolarizability molecules via joining the promising organic fragments: a DFT study, J. Mol. Model. 29 (2023) 1–11.
- [12] H. A. Kurtz, J. J. P. Stewart, K. M. Dieter, Calculations of the nonlinear optical properties of molecules, J. Comput. Chem. 11 (1990) 82–87.
- [13] H. A. Kurtz, D. S. Dudis, Quantum mechanical methods for predicting nonlinear optical properties, in: K. B. Lipkowitz, D. B. Boyd (Eds.), *Reviews in Computational Chemistry*, Wiley, New York, 1998, pp. 241–279.
- [14] N. S. Labidi, A. Djebaili, I. Rouina, Substitution effects on the polarizability and first hyperpolarizability of all-trans hexatriene, J. Saudi Chem. Soc. 15 (2011) 29–37.
- [15] S. R. Marder, D. N. Beratan, L. T. Cheng, Approaches for optimizing the first electronic hyperpolarizability of conjugated organic molecules, *Science* 252 (1991) 103–106.
- [16] S. R. Marder, J. W. Perry, Nonlinear optical polymers: Discovery to market in 10 years?, *Science* 263 (1994) 1706–1707.
- [17] N. Metropolis, A. W. Rosenbluth, A. H. Teller, E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. 21 (1953) 1087–1092.
- [18] N. M. O'Boyle, M. Banck, C. A James, C. Morley, T. Vandermeersch, G. R. Hutchison, Open Babel: An open chemical toolbox, *J. Cheminf.* 3 (2011) 1–14.
- [19] M. B. Ros, Organic materials for nonlinear optics, in: J. Novga, D. Braga, L. Addadi (Eds.), *Engineering of Crystalline Materials Properties*, Springer, Dordrecht, 2008, pp. 375–390.
- [20] K. D. Singer, J. E. Sohn, L. A. King, H. M. Gordon, H. E. Katz, C. W. Dirk, Second-order nonlinear-optical properties of donor- and acceptor-substituted aromatic compounds, J. Opt. Soc. Am. B 6 (1989) 1339–1350.
- [21] J. J. P. Stewart, MOPAC: A semiempirical molecular orbital program, J. Comput. Aided Mol. Design 4 (1990) 1–103.
- [22] J. J. P. Stewart, Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements, J. Mol. Model. 13 (2007) 1173–1213.

- [23] P. J. M. Van Laarhoven, E. H. L. Aarts, Simulated Annealing: Theory and Applications, Kluwer, Dordrecht, 1988.
- [24] D. Weininger, SMILES, a chemical language and information system.
 1. Introduction to methodology and encoding rules, J. Chem. Inf. Comput. Sci. 28 (1988) 31–36.