https://doi.org/10.1016/j.jmgm.2020.107798

# Robust design of D-π-A model compounds using digital structures for organic DSSC applications

Feng Wang[1,2]*, Steven Langford[1] and Hiromi Nakai[3,4]

[1]Department of Chemistry and Biotechnology,

[2]Centre for Translational Atomaterials

Faculty of Science, Engineering and Technology,

Swinburne University of Technology, Melbourne, Vic. 3122, Australia

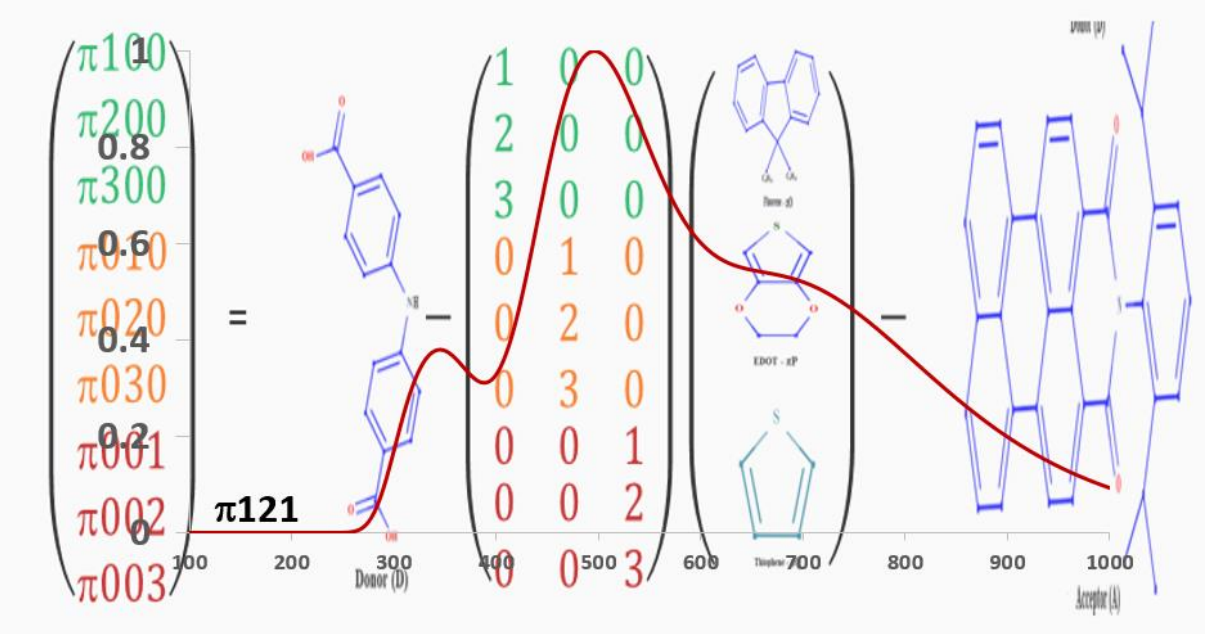
[3]Department of Chemistry and Biochemistry,

[4]Waseda Research Institute for Science and Engineering,

Waseda University, 3-4-1 Okubo, Shinjuku-ku,

Tokyo 169-8555, Japan

*fwang@swin.edu.au



**Abstract**

New D–π-spacer–A model compound dye sensitizers (or dyes) are developed using digital structures for organic dyes sensitized solar cells (DSSCs) applications. Based on our previous studies, the model D–π-spacer–A dyes contain building blocks of a di(p-carboxy)-phenylamine as the electron donor and a perylene monoimide as the electron acceptor. The new D–π-spacer–A dyes are constructed through variations of a set of three model π-spacer units, fluorene, 3,4-ethylenedioxythiophene and thiophene. The new dyes are presented by digital structures of π(ijk) in a digital control Π-matrix. If the chromophore database of the π-units is arranged in a defined manner, the new dyes are therefore designed through selecting of the set of three integers (ijk). Properties such as the UV-vis spectra which are calculated using the time-dependent density functional theory (TD-DFT) determine if the new compounds are suitable for organic solar cell purposes. The same strategy can be applied to donors and acceptors in the D–π-spacer–A model compound in order to robust design and build new organic dyes for DSSCs. The digital structures of the organic compounds enhance the machine driven structure-property relationship establishment once the database is sufficiently comprehensive. The present study demonstrates that new compounds obtained through mixing the π-spacer units of fluorene, 3,4-ethylenedioxythiophene and thiophene, e.g. π121 and π211, result in better dyes in DSSC applications. The concise digital structures of the new dyes are able to achieve a more robust design of the organic dyes and other materials.

**Key words:** Organic dye digital structures, Organic dye sensitised solar cells (DSSCs), Donor–π-spacer–Acceptor model, Structure-property relationship, UV-vis spectrum, Time-dependent density functional theory (TD-DFT) calculations, Feature engineering, Surrogate model for new organic dye design

## Introduction

Energy represents a major global challenge. Solar energy remains the most abundant renewable energy resource available to the earth, and must ultimately be the cheapest and most portable source of energy globally. In addition to all advantages of renewable energies, photovoltaic (PV) solar cells have the ability to make the current electricity infrastructure redundant. Solar cells can do this by bypassing the aging and fragile electricity grid to deliver power directly to the end user, fundamentally changing the underlying economics of energy and environment.[1] Among available solar cell technologies, organic dye sensitized solar cells (DSSC) are considered one of the most promising PV technologies. This is because their advantages of compatibility with simple, low-cost and high volume production methods, their optical and mechanical properties, unique flexibility and the high indoor efficiency.[2] [3]

At the heart of this technology, organic dyes (or OPVs) represent one of the most important components of an organic DSSC device, but have suffered long term by their low power conversion efficiency (PCE) since their invention in 1991.[4] The development of OPVs has long lingered at very low convert rate of the sunlight, until the recent breakthrough for the record PCE of 17.3%.[5] In this study, Meng et al. [5]were guided by a semiempirical model analysis and took advantage of the high diversity and easily tunable band structure of organic materials. The significantly improvement in efficiency of OPVs has become an engine to boost development of high performance OPVs through rational design both experimentally and theoretically.[6] However, the best PCE is still much lower than the Shockley–Queisser (SQ) limit of ca ~34%. [7] As a result, to design or to discover new OPVs that can absorb a broad electromagnetic spectrum, and subsequently converts to electrical work is a key development in solar energy.

A paradigm shift is required in design of organic dyes (OPVs). The properties of D-π-A model OPVs can be tuned by varying the donor, the π-spacer, and the acceptor moieties, respectively. A large number of conjugate organic molecules are suitable organic dye building block candidates. Databases such as the Harvard clean energy project contain approximately 2.6 million candidate compounds with their initial ranking.[8] Obviously, these organic dye building block candidates will construct a huge number of possible OPVs. Due to combinational complexity, it is not possible to synthesize or even to calculate such a large number of potential OPVs, even with application of a number of empirical selection rules. Current computer aided rational design for organic DSSCs is intuitive and less robust. The trial and error approach in OPV design estimates the light

absorption from generation of the UV-Vis spectra of compounds obtained from chemical modifications of an existing high performance dye,[9] [10] [11, 12] can be a very expensive and arbitary process, as chemical intuation does not always work. The strategy of such chemical modification, which targets the optimal absorption efficiency occurs between 680–700 nm[13]. If computer rational design of OPVs can be assisted by machine learning (ML) or artificial intelligence (AI), the OPV design will become particularly powerful when integrating with organic moiety databases to design more efficient OPVs. Before entering into ML and AI phase of development, the bottleneck problem is feature engineering of the database. It includes to discover a concise and efficient structure-property relationship of the OPVs which computers can recognise easily. As a result, training the algorithm needs well designed and featured engineered initial dataset and the present study aims at production such the D-π-A OPV dataset for this purpose. In the present study, we developed the integer digital structure code for the OPV development.

## Method and computational details

In the present study, following the development of digital structures in our previous work,[11] the model compounds are based on the same class of high performance D–π spacer–A model or template OPVs.[14, 15] This class of OPVs employed di(p-carboxy)-phenylamine (DCPA) as the electron donor and a perylene monoimide (PMID) as the electron acceptor, which are excellent motifs to connect to the electrodes [16, 17] in DSSC applications.[14, 15, 18] Theoretically, one can fix any two building blocks among the donors (D), acceptors (A) and π spacers from organic DSSC databases, to produce new dyes. This study extends to more combinations among the π spacer building units of fluorene (FLN), 3,4-ethylenedioxythiophene (EDOT) and thiophene (THN) to produce new dyes. The chemical structures of the building blocks are given in Figure 1, which is able to produce necessary results in the database to train suitable algorithms in ML. In this study, we explore the relationship between UV-vis spectra and the photophysical properties of the new dyes of different π-spacers through various combinations of FLN, EDOT and THN.

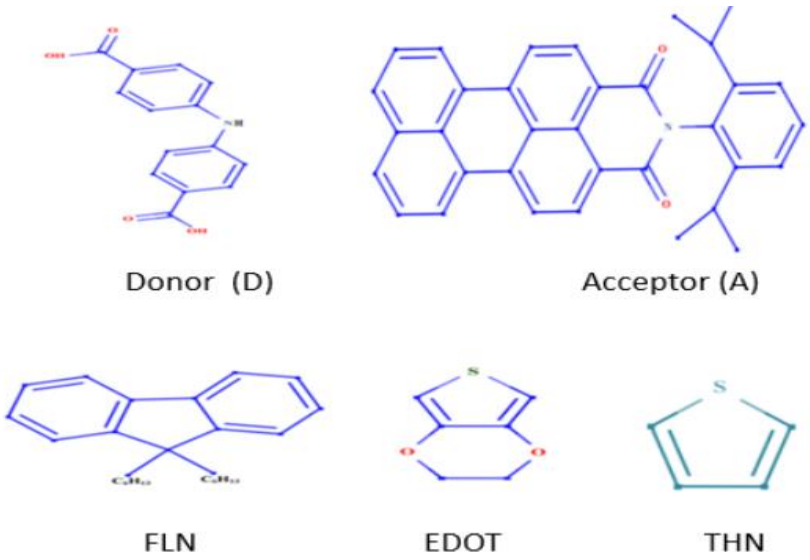



**Figure 1.** Chemical structure of the new organic dye fragment units of the D–π-spacer–A dyes. The donor (D), acceptor (A) and π-spacer units (FLN, EDOT and THN). The π-spacer contains only three units (i.e., FLN, EDOT and THN) in the present study and can be extended to many.

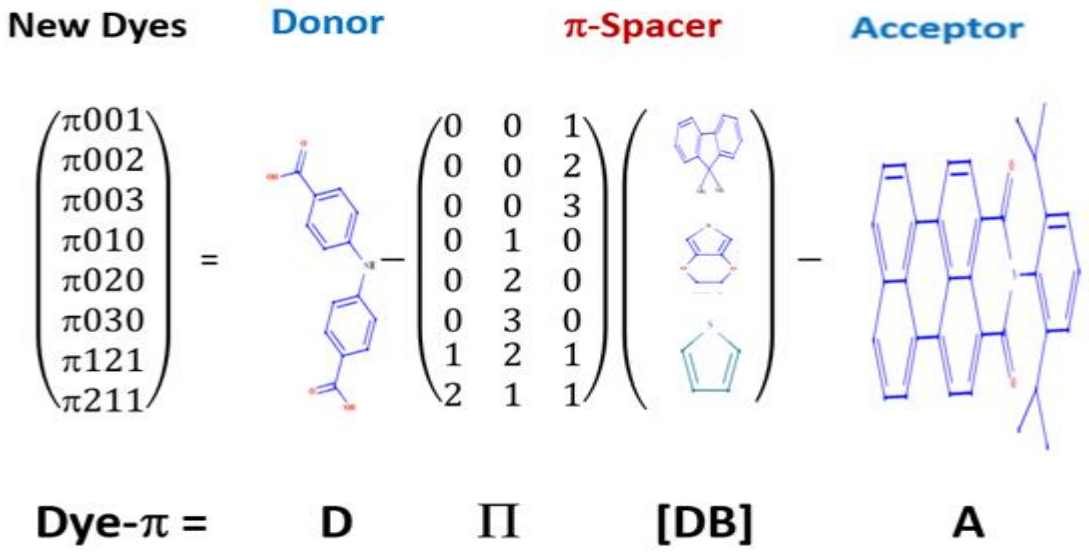
The computational details to calculate the new dyes are the same as previous study for comparison and consistency purposes in the same dataset.[19] Specifically, the new organic dyes are made from the same building blocks, i.e., the donor, DCPA, the acceptor, PMID and the π-spacer units, FLN, EDOT and THN (Figure 1) through various combinations of the π-spacer units. The building blocks were saturated using hydrogen atoms as monomers in the calculations. Note that only position isomers not conformers are considered as different dyes in the present study for simplicity. There is a large number of available density functional theory (DFT) functionals for time-dependent calculations as listed in the DFT popularity poll [20] and some is more populate than the others. However, the quality of the DFT functional prediction is largely dependent on the property, the molecular system, environment (gas phase or solution) and basis set employed [21, 22]. Sini et al [22] assessed over a dozen of DFT functionals for tetrathiafulvalene-tetracyanoquinodimethane model complex in the cofacial geometry, it was discovered that the ground-state charge transfer varied in a continuous manner as a function of the (donor) HOMO-(acceptor) LUMO energy gap[22]. As a result, perhaps a valid and reliable method for a reliable DFT functional in the calculations is the agreement with the measurement of the property. Our previous UV-vis spectral calculations for this class of compounds have been proven in good agreement with experimental measurements.[14, 15, 18, 19] As a result, the DFT based PBEPBE/6-31G(d) method was employed to re-optimize the new dyes in tetrahydrofuran (THF) solution and for the UV-vis spectral calculations, from the optimized geometries obtained using

the B3LYP/6-31G(d) model in isolation. No imaginary frequencies were discovered for the optimized new dyes, indicating that the obtained structures are stable at a true energy minimum. The calculations in THF solvent were based on the polarizable continuum model (PCM). The UV-vis spectra in THF solution were calculated for singlet-singlet transitions using the time-dependent (TD)-DFT method up to the 30$^{th}$ lowest spin-allowed excited states for each dye structure. All calculations were carried out Gaussian16 computational chemistry package.[23]

## Results and discussion

### *New OPV digital structure π(ijk) and the Π-matrix*

OPVs represent one of the most important elements of a DSSC device.[2] To design high performance new dyes based on the pull-push (D-π-A) configuration, the present study keeps the same donor (DCPA) and acceptor (PMID) but the new dyes are constructed with different combinations of FLN, EDOT and THN for π-spacer. In order to assess the structure-spectrum relationship of the DCPA-π-PMID new dyes with respect to systematically change of the π units, we designed new dyes in the form of a surrogate model as DCPA-[$(FLN)_i$-$(EDOT)_j$-$(THN)_k$]-PMID. For simplicity, if the chemical connection order of the π-spacer, -[$(FLN)_i$-$(EDOT)_j$-$(THN)_k$]- is not interchangeable and the integers i, j and k = 0,1,2,3 and $i+j+k \leq 4$ to restrict the length of the π-spacer ($L_\pi$) for being up to four unit long, it results in 32 possible π-spacers and therefore 32 possible new organic dyes. Once the donor (DCPA) and acceptor (PMID) are determined, the new dyes are produced uniquely through the elements (i j k) of the control n×3 Π-matrix in Scheme 1 by Arooj and Wang[19].



**Scheme 1.** Design of new πijk organic dyes using the DCPA-Π-matrix [π-spacer Database]-PMID template and digital structures. The [π-spacer Database] in this study contains FLN, EDOT and THN units which are controlled by the Π-matrix by Arooj and Wang[19].

The digital OPV structure π121 defined in Scheme 1 extracts the OPV structural code into a control Π-matrix.[19] Here in the *n*×3 order Π-matrix, n is the number of new dyes designed and 3 is the number of π-units, FLN, EDOT and THN employed in this study. This *n*×3 order Π-matrix controls the new OPVs in the n×1order **Dye-π** matrix in Scheme 1 (in this case *n*=8 which is the number of new dyes to be constructed) through the *n*×*m* (*m*=3) control **Π-matrix** with integer matrix elements. The number of the π-spacer units in the database **[DB]** (*m*×1, in this case *m*=3). Once the surrogate model is defined, the π-spacer database **[DB]** can be plugged into other organic chromophore databases with more than three π-spacer units. For example, a new OPV with the digital structure of π121 containing 1 FLN; 2 EDOT and 1 THN is coded as (121) in the Π-matrix and the chemical structure of π121 is written as, DCPA-FLN-EDOT-EDOT-THN-PMID in Figure 2. The π-spacers of the eight new DCPA-π-PMID OPVs given in Figure 2 are obtained by setting up the elements of the Π-matrix in Scheme 1 to (010), (020), (030), (001), (002), (003), (121) and (211) (only the π-spacer units are given). The digital structures can be considered as an analogy to the quantum mechanical basis sets, such as Pople style 6-31G basis set.

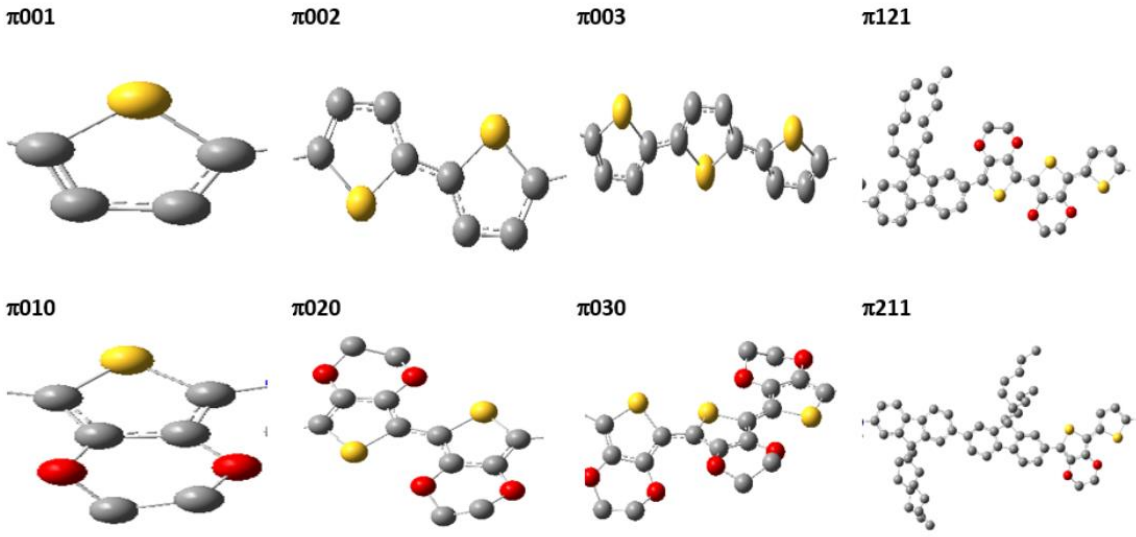



**Figure 2.** The digital and optimised chemical structures of the π-spacers of the eight new DCPA-π-PMID OPVs in THF solvent (only the π-spacers are presented as the donor and acceptor are the same for all dyes).

### *Tuning HOMO-LUMO energy gap of the new OPVs using πijk*

The properties of the structures of the new OPVs in Figure 2, which are produced using Scheme 1 are calculated using DFT method detailed in previous section. The target of the new OPV design is (a) to turn the absorption (hyperchromic) valley formed by the major UV-vis spectral bands between the DCPA donor and the PMID acceptor in the UV-vis spectra[19] into a band, and (b) to shift the optimal absorption band towards the region of 680–700 nm.[13] That is, the optical absorption should cover a large part of the visible spectrum and extend up to the NIR region; high molar extinction coefficients are desired to achieve efficient solar photons harvesting.[2] This can be done through rationally changing/modifying the chemical structures of existing OPVs in a trial and error approach. Table 1 summarises the molecular properties of the new OPVs (top) together with the same class of OPVs obtained in previous studies.[14, 15, 19], calculated using the same DFT methods.

**Table 1.** Selected molecular properties of the related organic dyes in their digital codes (πijk)*.

| OPVs | Mol. Formula | $\mu$ (D) | $L_\pi$[a] (A°) | $<R^2>$ (a.u) | $\Delta\varepsilon$[b] (eV) | $I/I_{max}$ |
|---|---|---|---|---|---|---|
| | | **Present Study** | | | | |
| **π010** | $C_{73}H_{68}N_2O_6$ | 8.95 | 5.31 | 99669.05 | 1.37 | 0.34 |
| **π020** | $C_{60}H_{44}N_2O_{10}S_2$ | 9.40 | 9.27 | 146440.56 | 1.24 | 0.45 |
| **π030** | $C_{66}H_{48}N_2O_{12}S_3$ | 13.25 | 12.75 | 206541.34 | 1.14 | 0.58 |
| **π001** | $C_{52}H_{38}N_2O_6S$ | 5.57 | 5.93 | 95769.82 | 1.32 | 0.36 |
| **π002** | $C_{104}H_{104}N_2O_8S$ | 2.15 | 9.22 | 142028.10 | 1.05 | 0.47 |
| **π003** | $C_{60}H_{42}N_2O_6S_3$ | 7.06 | 13.14 | 199681.91 | 0.92 | 0.64 |
| **π121** | $C_{129}H_{136}N_2O_8S$ | 7.07 | 21.45 | 692844.94 | 2.11 | 1.0 |
| **π211** | $C_{108}H_{106}N_2O_8S_2$ | 8.81 | 25.29 | 416579.57 | 1.10 | 0.88 |
| | | **Previous Study** | | | | |
| **π100** | $C_{73}H_{68}N_2O_6$ | 5.12 | 6.90 | 165396.79 | 2.70 | - |
| **π200** | $C_{123}H_{132}N_2O_6$ | 9.85 | 23.51 | 588627.82 | 2.71 | - |
| **π300** | $C_{104}H_{104}N_2O_8S$ | 8.99 | 18.98 | 429017.70 | 2.68 | - |
| **π110** | $C_{79}H_{72}N_2O_8$ | 4.83 | 10.86 | 234059.85 | 2.54 | - |
| **π111** | $C_{83}H_{74}N_2O_8S_2$ | 8.31 | 14.71 | 301206.28 | 2.35 | - |
| **π210** | $C_{104}H_{104}N_2O_8S$ | 6.76 | 19.17 | 435121.75 | 2.52 | - |
| **π310** | $C_{129}H_{136}N_2O_8S$ | 8.56 | 27.18 | 692844.94 | 2.48 | - |
| **π220** | $C_{110}H_{108}N_2O_{10}S_2$ | 15.93 | 22.46 | 520845.22 | 2.30 | - |

*In THF solution.

[a]$L_\pi$ denotes the length of "Linker Bridge" in a D-π-A OPV.[9] That is, $L_\pi$ represents the length between the nitrogen atom of the donor (DCPA) and the carbon atom of the acceptor (PMID).[19] The present lengths of "Linker Bridge" of the OPVs are not exactly the same with the previous dyes in Ref [19] as the former defined as carbon to carbon.

[b]The HOMO-LUMO energy gap of the OPVs.

From energy point of view, the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) is the most energy favourable transition in UV-vis spectra as it is the smallest energy required for a UV-vis transition. Figure 3 reports the HOMO-LUMO energy gaps (eV) of the new OPVs (green) and the previously designed OPVs (blue). Apparently, the HOMO-LUMO energy gaps of the OPVs in previous study,[19] i.e., the blue bars are larger than the green bars of the OPVs in the present study. This trend indicates that EDOT and THN π-units contribute to the reduction of the HOMO-LUMO energy gap more efficiently than the FLN π-unit alone and their combinations. That is, the π0x0 and π00x OPVs in Figure 3 exhibit smaller HOMO-LUMO gaps (Δε) than the πx00 OPVs in the same Figure. For example, the HOMO-LUMO gaps of the π0x0 and π00x OPVs are less than 1.5 eV without the FLN π-unit, whereas this energy gap of the πx00 OPVs [19] are all larger than 2.0 eV in the Figure (also see second last column of Table 1).

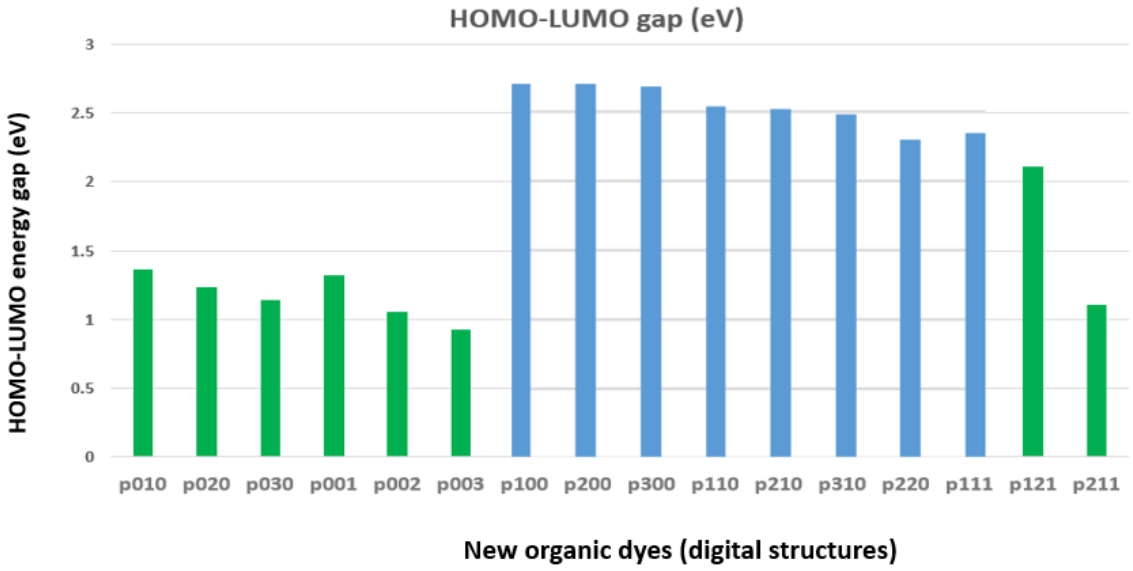



**Figure 3.** The HOMO-LUMO energy gaps (eV) of the new organic dyes (green) and the previously designed organic dyes (blue). The data are from the second last column of Table 1.

Tandem of π-units does not necessarily reduce the HOMO-LUMO energy gap of the OPVs as shown in Figure 3, as the chemical structures and interaction among the building blocks of the OPVs follow quantum mechanics. For example, increase of the number of FLN building blocks in the D-π-A OPVs in π100, π200 and π300 does not apparently change the HOMO-LUMO energy gap (Δε) as indicated by Arooj and Wang[19]. The HOMO-LUMO energy gap (Δε) of π100, π200

and π300 OPVs in Figure 3 (also see Table 1), i.e., DCPA-FLN-PMID, DCPA-FLN-FLN-PMID and DCPA-FLN-FLN-FLN-PMID, are 2.70, 2.71, and 2.68 eV, respectively.[19] However, tandem of the EDOT and THN π-units indeed, reduce the HOMO-LUMO energy gap (Δε). For example, π001, π002 and π003 OPVs in Figure 3, DCPA-THN-PMID, DCPA-THN-THN-PMID and DCPA-THN-THN-THN-PMID, are 1.32, 1.05, and 0.92 eV, respectively. The HOMO-LUMO energy reduction will quickly flat out and the "magic number" usually is no more than four.[24] If the OPV contains only single π-units, the HOMO-LUMO energy gaps reduction is less for tandem fluorene (πx00) only OPVs than for tandem EDOT only π0x0 and or tandem THN only π00x OPVs.[19] Figure 3 also demonstrates that certain combination of the three π-units such as π211, i.e., DCPA-π(FLN)-π(FLN)-π(EDOT)-π(THN)-PMID significantly reduces the HOMO-LUMO energy gap (Δε) of the new OPV to 1.10 eV. The digital OPV structure helps the pattern recognition of the OPVs.

### *Tuning the UV-vis spectra of the new OPVs using πijk*

The simulated UV-vis spectra in the THF solvent of the eight new OPVs in Figure 2 are given in Figure 4. The UV-vis spectra of the π0x0 and π00x OPVs such as π010, π020 and π030 (dashed spectra in Figure 4) and π001, π002 and π003 (dot spectra in Figure 4) exhibit a very different shape comparing to previous πx00 OPVs in which the FLN unit(s) made the π-spacers.[19] The number of the FLN in the πx00 OPVs enhances the UV-vis spectra in the major band region of 300-450 nm, whereas the impact to the target band of 450-600 nm of the spectra is small and can be ignored.[19] The UV-vis spectra of the new OPVs (π0x0 and π00x) in Figure 4 show that the positions of the major spectral bands are about 600 nm with a broad coverage in 300-1000 nm, which is within the region of 680-800 nm for the target OPVs.[13] Hopefully, these OPVs have the potential to be more efficient in the DSSC applications. If the $TiO_2$ and $I_3^-/I^-$ pair is employed for the DSSC device, the HOMO and LUMO energy levels of the sensitizers (OPVs) need to be out side of -4.92 eV ($I_3^-/I^-$) and -4.05 eV ($TiO_2$), respectively. For these reasons, only π121 sensitizer meets this condition. [2, 9]

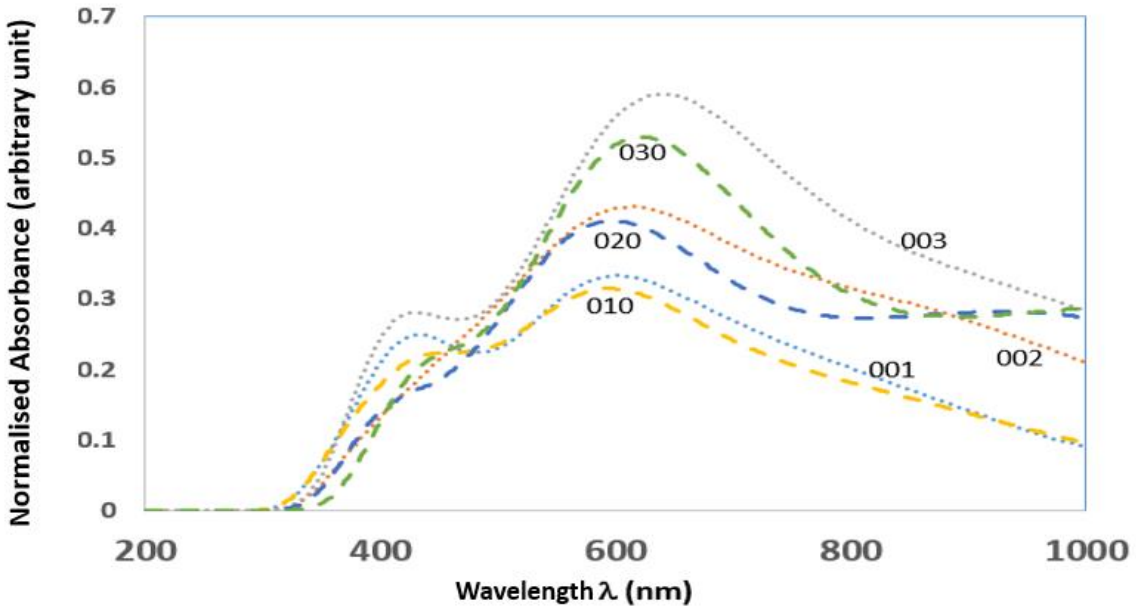



**Figure 4.** The simulated UV-vis spectra in THF solvent of the new π0x0 (dashed lines) and π00x OPVs (dot lines). All spectra are normalized to $I_{max}$ of new dye π121 (see Table 1).

Directly comparing the calculated UV-vis spectra of the new OPVs, it reveals that contributions of the EDOT (π0x0) π-spacers and THN (π00x) π-spacers of the OPVs lead to the preferred hyperchromic spectral shift in a similar degree with THN spacer OPVs (π00x) slightly better performance than the EDOT-spacer OPVs (π0x0). Both of the new dyes with π0x0 and π00x π-spacer OPVs perform better than the fluorene spacer only OPVs (πx00).[14, 15, 19]

Figure 5 compares the UV-vis spectra of the OPVs with mix π-spacer π121 and π211 with respect to the OPVs with only single π-spacers, that is, FLN π100, EDOT π010 and THN π001 OPVs. As found by the previous study[19] that the major spectral band pattern of the donor monomer (DCPA) at 330 nm, and the acceptor monomer (PMID) at 500 nm. The OPVs (πx00) with FLN only π-spacers, i.e., π100, π200 and π300, produce a UV-vis band at approximately 400 nm, which is the location of the "valley" formed by the donor monomer and the acceptor monomer. In other words, the OPVs (πx00) with FLN only π-spacers indeed contribute to red shift the UV-vis band of the donor (DCPA) monomer from 330 to 370 nm and also contribute less significantly to red-shift (20-30 nm) the spectral band from the acceptor (PMID). [19]Mixing the FLN (π100) and EDOT (π010) units in the π-spacers such as (π110), (π210) and (π310) red-shift the OPVs (πx00) from 370/400 to 400/600 nm, without significant spectrum changes. [19]

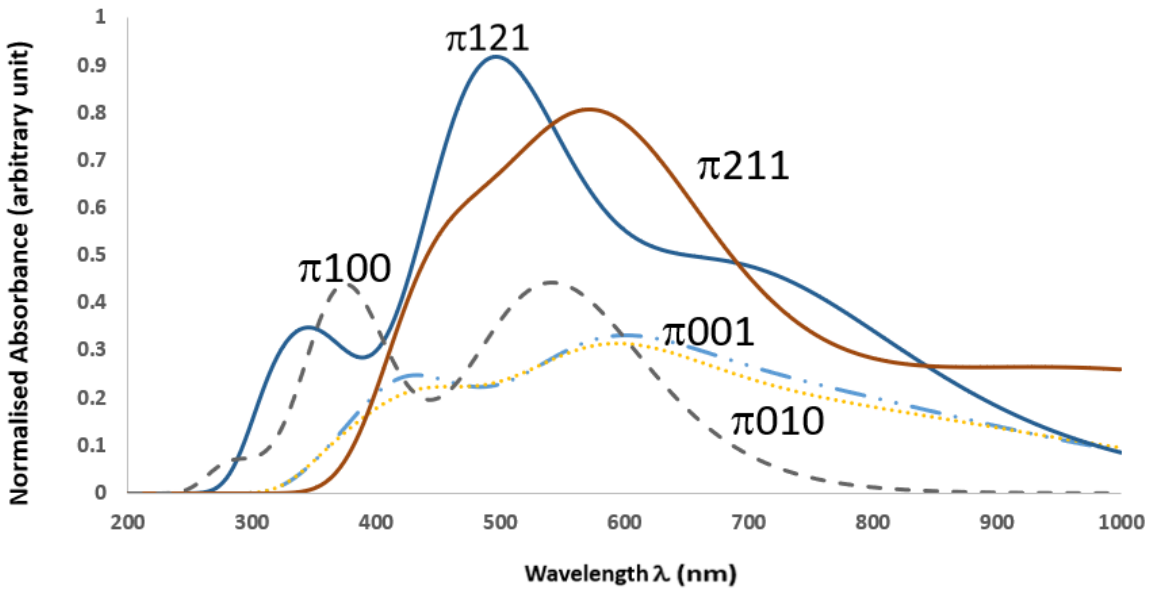



**Figure 5.** Simulated UV-vis spectra of the best performing dyes of π121 and π211 in this study. Together with the spectra of π100, π010 and π001 dyes with only one π unit of the π-spacer (in the same THF solvent).

Mixing of the three π-spacer units of FLN, EDOT and THN, which can be achieved varying the elements of the control Π-matrix in Scheme 1. The new OPVs produce even more preferred UV-vis spectra simulated in the same condition. Figure 5 gives the UV-vis spectra of π121 (solid blue spectrum) and π211 (solid red spectrum), together with π100, π010 and π001 OPVs for comparison, in order to obtained information of the π-spacers. As seen in this figure, the spectra OPVs π121 and π211 are very promising. Addition of either a FLN into the π-spacer of π111 OPV or a THN into the π spacer of the π210 OPV, will produce π211 OPV with further enhancement of the bathochromic (red) shift to the wanted direction.

Table 2 summarizes the transitions of the major spectral band positions and intensities of the new OPVs. The spectral bands with the oscillator strength $f > 0.20$ are collected in this table and the transitions with >10% are collected. The most intensive band ($f_{max}$) of the UV-vis spectra of the OPVs are highlighted and the band due to the HOMO (H) → LUMO (L) transitions of the OPVs are coloured in blue in the table. The highlighted and blue coloured transitions are not always the same for an OPV in this table. All the OPVs possess $C_1$ point group symmetry so that the symmetry forbidden transitions are relaxed. As a result, the dominant transitions to the UV-vis bands in this table populate among the frontier orbitals of the OPVs, covering from (H-4) to (L+5). For example, the most intensive band of the π010 OPV at $\lambda$ = 588 nm with the spectroscopic strength of 0.47 is

dominated by transitions of (H-1) → L (21%) and (H-2) →L (21%), whereas the next strong band of this same OPV at $\lambda$ = 840 nm ($f$ = 0.32) is dominated by the H → L (41%) transition. However, the most intensive band of the π030 OPV at $\lambda$ = 1122 nm with the spectroscopic strength of 0.66 is dominated by the transition of H → L (48%). However, this transition is out of the UV-vis region and into the IR region. In addition, there is no strong UV-vis spectral bands of π020 OPV dominant by the H → L transition.

**Table 2.** Calculated dominate UV-vis transitions λ (in nm) of the new OPVs in THF solvent*.

| OPVs | λ (f>0.20) | Dominant transitions (>10%) | |
|---|---|---|---|
| π010 | **588.28(0.47)** | (H-1) → L | (21%) |
| | | (H-2) → L | (20%) |
| | 839.58(0.32) | H → L | (41%) |
| | 435.48(0.25) | (H-2) → L | (29%) |
| | | (H-1) → L | (17%) |
| | 628.85(0.22) | (H-1) → (L+1) | (21%) |
| π020 | **987.27 (0.63)** | (H-1) → L | (47%) |
| | | (H-2) → L | (13%) |
| | 669.31 (0.35) | (H-1) → (L+1) | (30%) |
| | | (H-2) → L | (13%) |
| | 584.53(0.24) | (H-3) → L | (30%) |
| π030 | **1122.62(0.66)** | H → L | (48%) |
| | 593.93(0.52) | (H-4) → L | (22%) |
| | | H → (L+2) | (12%) |
| | 640.43(0.41) | (H-1) → L | (21%) |
| | 741.06(0.27) | (H-1) → L | (32%) |
| | | (H-2) → L | (15%) |
| π001 | **594.87(0.71)** | (H-1) → L | (36%) |
| | | H → L | (11%) |
| | 802.66(0.38) | H → L | (39%) |
| | | (H-1) → L | (11%) |
| | 425.33(0.32) | (H-1) → (L+1) | (41%) |
| π002 | **628.94(0.80)** | (H-1) → L | (38%) |
| | 883.38(0.58) | H → L | (43%) |
| π003 | **685.70(0.70)** | (H-1) → L | (32%) |
| | | (H-2) → L | (12%) |
| | 613.24(0.66) | (H-2) → L | (23%) |
| | | (H-2) → (L+2) | (19%) |
| | 970.03(0.60) | H → L | (46%) |
| π121 | **476.63(1.31)** | (H-1) → L | (40%) |
| | 521.29(1.07) | (H-1) → L | (37%) |
| | 709.50(1.06) | H → L | (48%) |
| π211 | 455.33(0.64) | (H-1) → (L+3) | (19%) |
| | **583.15(1.19)** | (H-3) → L | (16%) |
| | | H → (L+1) | (16%) |
| | 1018.35(0.55) | H → (L+5) | (10%) |
| | | H → L | (47%) |

*Here H-1 represents for the next HOMO and L+1 represents the orbital above the LUMO.

The results in Table 2 reveals that the UV-vis spectrum of an OPV depends not only the HOMO-LUMO energy gap, but also the frontier orbitals and their distribution---the (allowed, i.e., singlet) low-lying excited states (L+m) and the occupied outer valence orbitals (HOMO-n). The transitions

between (H-*n*) → (L+*m*) (where n and m are integers) can be more probable than the HOMO-LUMO transition for larger OPVs such as these in this study. Therefore, in Table 2 the dominant UV-vis transitions of the OPVs depend on the HOMO-LUMO energy gap as well as the energies of hole states (the singly occupied states) and the excited states where $n$ = {0, 1, 2 and 4} and $m$ = {0, 1, 2, 3, 4 and 5}, here $n$ = 0 for HOMO and $m$ = 0 for LUMO. Of the eight new OPVs in Table 2, none of the strongest transitions of the OPVs are dominated by the HOMO-LUMO transitions within the target region of 400-800 nm. Figure 6 reports the UV-vis spectrum of the π121 OPV and the frontier orbitals whose transitions dominate the major bands. The dominant transitions, which are (H-1) → L and H → L, also indicated in the spectrum. Nevertheless, the HOMO-LUMO energy gap of an OPV is a useful indicator for the minimum possible energy required, it is however, not necessarily the most reliable indicator for the most intensive band of the UV-vis spectra as many other allowed transitions of the OPVs exhibit preferred spectral shifts.

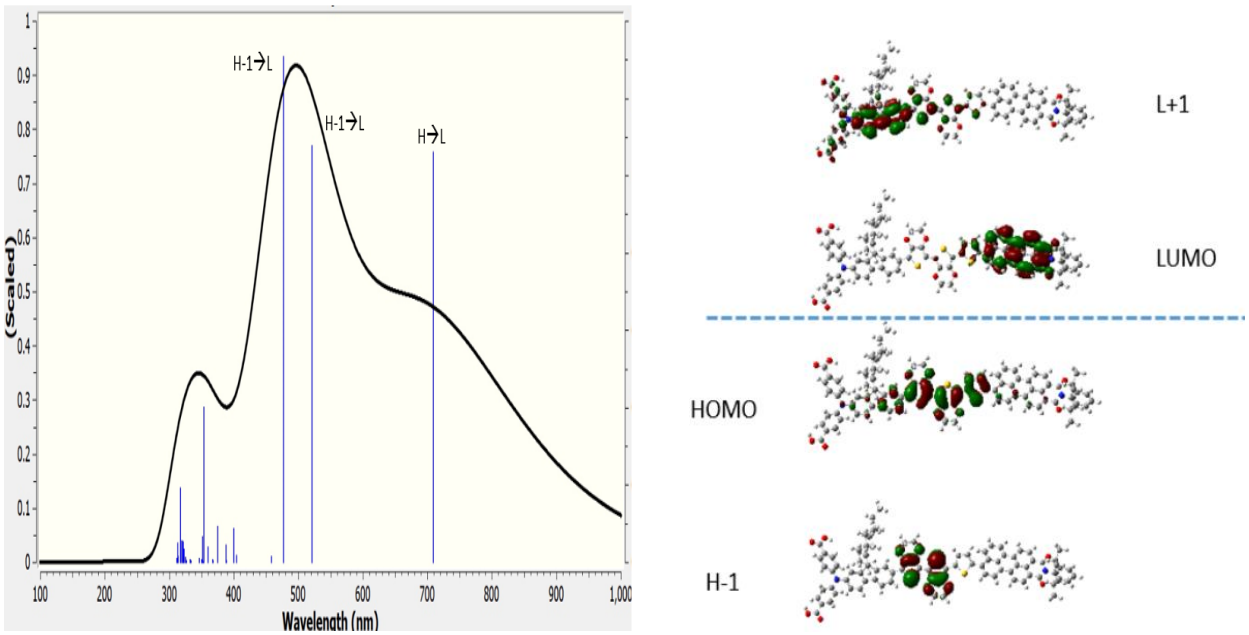



**Figure 6.** Simulated UV-vis spectrum of the π121 organic dye (left hand side). The frontier orbitals of this compound which dominate the UV-vis transitions, are given on the right hand side.

Figure 6 also plots the frontier orbitals of H-1 (376a), HOMO (377a), LUMO (378a) and L+1 (379a) of π121 OPV. The HOMO density concentrates on the π-spacer, part of FLN and two EDOTs and THN, the LUMO concentrates on the acceptor (PMID). The density of HOMO-1 concentrates on the two EDOT unit of the π-spacer but the LUMO+1 concentrates on FLN next to the donor. The most intensive transition of the π121 OPV is (H-1)→ L.

## Robust design and turn new organic materials into digital permutation

To design state-of-the-art materials such as high-performance organic materials, one needs the paradigm shift. Experimental synthesis provides valuable insight into the general structure of chromophores, but the ability to obtain information at the atomic level in a suitable timeframe is limited.[25] Computer aided design accurately predicts the properties of the given (often from synthesis) material structures, and avoids the costly, lengthy and inefficient empirical design-synthesis-fabrication-testing cycle. In Scheme 1, the donor and acceptor of the organic materials are fixed and only the π-spacer in the D-π-spacer-A structure is allowed to change among upto three ($n \leq 3$) building units in the new dye design. This restriction leads to up to 32 new organic dyes, under the conditions that no more than four π-units for the length of the π-spacer ($N_\pi \leq 4$) and no permutation for the π-spacer units.

A large number of conjugate organic chromophores are suitable for building blocks for the organic dye purposes in solar cell applications. If the chromophores database [DB] in Scheme 1 is replaced by, for example, the small dataset of 27 π-spacer units,[26] it could build hundreds of thousands of possible organic dye candidates like the 26 letters in the alphabet for a dictionary. In addition, this same strategy can be applied to donor and acceptor for new high performance organic dyes. The combinational complexity will produce a significant number of possible organic dye candidates, which quickly exceed the human capability of measurements and computational resources. As a result, ML is the direction in the development of a robust material design for suggest suitable structures for solar cell applications.

A paradigm shift is needed in organic dye design. The structural information of the organic chromophores at the atomic orbital level is central to further improvement and rational design of new solar cell devices. A conventional organic dye development starts from input (A, the experimentally obtained high performance dyes) with human expert provided discovery rules (B, the structure-property relationship) to obtain output (C, the new dyes), i.e. A+B→C. It applies to both a trial-and-error based empirical design approaches through direct synthesis [14, 15, 18] and an insightful but assumption based computer simulation such as modification of recognised high performance organic dyes (Arooj and Wang, 2019). In an inverse organic dye design, however, the discovery rules (B) are developed from input (A, structure) and output (C, target properties) of both existing experiment and quantum mechanical calculations in a "learning to learn" manner, i.e., A+C→B. The input (A) and output (C), which are obtained from conventional approaches to

form the training dataset to develop the discovery rules (B). Clearly, ML is the technique to be employed to design new organic dyes (A') with target properties C. As a result, an inverse design directly targets the properties of the right structure of organic dyes.

The bottleneck for the development of new materials using ML is to develop and to feature engineer the initial dataset. Although conventional organic dye design is less efficient, it is necessary as it provides the initial dataset to train the ML algorithms. The initial dataset is feature engineered to extract features (domain knowledge or properties) from raw data via data mining techniques. In the present design of D-π-A organic dyes, the advantages of feature engineering in Scheme 1 are as follows. First, it turns chemical structural information to concise digital information coded in the control Π-matrix at a more abstract level, once the π-spacer units are given in the database [DB]. The design of a new chemical structure of an organic dye becomes turning on and off the elements (i, j, k ≤ $N_\pi$ are integers where $N_\pi \leq 4$ in the present study) in the control Π-matrix. Variations of the integers in (ijk) in the control Π-matrix produce new organic dyes with clear structure features (given patterns). Second, Scheme 1 more efficiently extracts the structure-spectrum relationship of the organic dyes, particularly the impact of the π-spacer units on the properties (e.g. UV-vis spectra) of the class of D-π-spacer-A organic dyes. This will greatly help the quantum chemical based design strategies for new organic dyes and the application of ML. For example, increase of the integer k from 1 to 3 in the digital structures of π00k organic dyes (i.e., DCPA-π$(EDOT)_n$-PMID where $n$ =1, 2 and 3), the spectral bands at approximately 600 nm of the UV-vis spectrum (in Figure 4) exhibit the preferred hyperchromic and bathochromic spectral shift.

Third, Scheme 1 is adoptive and can be treated as a surrogate model to larger database. The database **[DB]** of the π-spacer units can be plugged in other database such as Guillen-Lopes et al. [27]. The n×3 order control Π matrix in Scheme 1 becomes $n \times m$ ($n$ is the number of new organic dyes and $m$ is the number of the conjugate π-units in the database) in the surrogate model. The same strategy is applicable to the donor (D) database or the acceptor (A) database, or three databases for donor, π-spacer and acceptor. Finally, Scheme 1 can be applied to design new organic compounds and other materials, such as catalysis, perovskites and drugs. Once trained, the surrogate models are able to reproduce (or converge) the original data with allowed accuracy (through a self-consistent field) and, more importantly, emulate certain properties such as UV-vis spectra of the

organic dyes in the present study produced by quantum mechanical calculations with significantly less resources.

## Conclusion

Design and synthesis of new organic dyes (and other chemicals) is a very expensive exercise. It is not practical even to screen quantum mechanically all D-π-A at design stage,[14, 15, 18, 19] when a large number of candidates are available. A paradigm shift is required in design of new materials. The present study proposed a surrogate digital structure model to systematically design new organic dyes D-π-A through integer permutation of the π-spacer units from the database for the DSSC applications. When combining quantum mechanical calculations with reclaiming from literature result, a database of organic dye building blocks can be constructed to prepare towards high-throughput ML techniques, which will be applied to suggest new high performance organic dye structures once we collect sufficient information in this regard.

In the present study, we presented a surrogate model to design high performance DCPA–π-spacer–PMID class of organic dyes through a control Π-matrix. It is discovered that when fixing the donor (DCPA) and the acceptor (PMID) in the DCPA–π-spacer–PMID organic dyes, repeating any one of the π-units, i.e., fluorene, EDOT and THN alone in the π-spacer is less effective than a mixture of these π-spacer, in agreement with our previous studies. [14, 15, 19]The UV-vis spectra of mixed organic dyes such as π121 and π211 exhibit preferred spectra among all the class of organic dyes studied so far. The present study further reveals that although the HOMO-LUMO energy gap is a good descriptor, the most intensive UV-vis band of the organic dyes is not necessarily the transition between the HOMO-LUMO, other frontier orbitals maybe involved. As a result, for the design of new high performance organic dyes, frontier orbitals and their distributions are important properties for a good organic material.

The present study showcases the great potential of the surrogate model to design high performance DCPA–π-spacer–PMID organic dyes using digital structure. It is an analogy to the Pople quantum mechanical basis set approach in which the code of a basis set such as 6-31G, represents a particular contract scheme of the gaussian primitives which are stored in the basis set database. The digital structures for the organic dyes paves the path future computer machine learning.

**Conflict of Interest**

There is no conflict of interest.

## Acknowledgment

FW acknowledges her research and project students, Ms Qudsia Arooj, Mr Joshua Ince and Mr Frederick Backler for contributions for quantum mechanical calculations of related projects. FW also acknowledges National Computational Infrastructure (NCI) and Swinburne Supercomputing Facilities for the supercomputing resources.

## References

1. Switching on optical properties of D-pi-A DSSC sensitizers from pi-spacers towards machine learning. Arooj, Q., Wang, F. 2019, Solar Energy, p. 1189.

2. Frisch, M. J., Trucks, G., Schlegel, H. B., Scuseria, G., Robb, M., Cheeseman, J., Scalmani, G., Barone, V., Mennucci, B., Petersson, G.,. Gaussian 09, Revision A. 1. Gaussian Inc. s.l. : Wallingford CT, 2016.

3. The Harvard Clean Energy Project: Large-Scale. Hachmann J., Olivares-Amaya R., Atahan-Evrenk S., Amador-Bedolla C., Sánchez-Carrera R.S., Gold-Parker A., Vogt L., Brockway A.M., Aspuru-Guzik A. 2011, Journal of Physical Chemistry Letters, p. 2241.

4. The First Picoseconds in Bacterial Photosynthesis—Ultrafast Electron Transfer for the Efficient Conversion of Light Energy. Zinth, W., Wachtveitl, J. 2005, ChemPhysChem, p. 871.

5. Rational use of ligand to shift the UV-Vis spectrum of Ru-complex sensitizer dyes for DSSC applications. Backler, F., Wilson,G., Wang, F. 2019, Radiation Physics and Chemistry , p. 66.

6. Determination of Trace Metals by Differential Pulse Voltammetry at Chitosan Modified Electrodes. Malavé Osuna, R., Capel Ferrón,C., Vercelli, B., Hernández, V., Zotti, G., López Navarrete, J.T.,. 2010, Portugaliae Electrochimica Acta, p. 63.

7. Electronic structure and nonlinear optical properties of organic photovoltaic systems with potential applications on solar cell devices: a DFT approach,. Guillén-López, A., Delesma, C., Amador-Bedolla, C., Robles, M., Muñiz, J. 2018, Theoretical Chemistry Accounts, p. 85.

8. Bradford, T. Solar Revolution: The Economic Transformation of the Global Energy Industry. s.l. : 1st Ed. MIT Press, 2008.

9. Research Progress on Photosensitizers for DSSC. Carella, A., Fabio Borbone, F., and Centore, R. 2018, Frontiers in Chemistry, p. 481.

10. Fine-tuning of the chemical structure of photoactive materials for highly efficient organic photovoltaics. Fan, B., Du, X., Liu, F., Zhong, W., Ying, L., Xie, R., Tang, X., An, K., Xin, J., Li, N., Ma, W., Brabec, C. J., Huang, F., Cao, Y. 2018, Nature Energy, p. 1051.

11. A low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO2 films. O'Regan, B., Grätzel, M. 1991, Nature, p. 737.
12. Organic and solution-processed tandem solar cells with 17.3% efficiency. Meng, L., Zhang, Y., Wan, X., Li, C., Zhang, X., Wang, Y., Ke, X., Xiao, Z., Ding, L., Xia, R., Yip, H.-L., Cao, Y., Chen, Y. 2018, Science, p. 1094.
13. NREL. National Renewable Energy Laboratory (NREL). [Online] 2020. https://www.nrel.gov/pv/cell-efficiency.html.
14. Developing photocathode materials for p-type dye- sensitized solar cells. Benazzi, E., Mallows, J., Summers, G.H., Black, F.A., Gibson, G.A. 2019, Journal of Materials Chemistry C, p. 10409.
15. Towards rational design of organic dye sensitizer solar cells (DSSC): an application to the TA-St-CA dye. Mohammadi, N., Mohan, P., Wang, F. 2013, Journal of Molecular Graphics and Modelling, p. 64.
16. Shifting UV-Vis absorption spectrum through rational structural modifications of zinc porphyrin photoactive compounds. Arooj, Q., Wilson, G., Wang, F. 2016, Royal Society of Chemistry (RSC) Advances, p. 15345.
17. Switching on optical properties of D-pi-A DSSC sensitizers from pi-spacers towards machine learning. Arooj, Q., Wang, F. 2019, Solar Energy, p. 1189.
18. Modulated charge injection in p-type dye-sensitized solar cells using fluorene-based light absorbers. Liu, Z.H., Xiong, D.H., Xu, X.X., Arooj, Q., Wang, H., Yin, L.Y., Li, W.H., Wu, H.Z., Zhao, Z.X., Chen, W.,Wang, M.K., Wang, F., Cheng, Y.B., He, H.S. 2014, ACS Applied Materials & Interfaces, p. 3448.
19. Fine tuning of fluorene-based dye structures for high efficiency p-type dye-sensitized solar cells. Liu, Z.H., Li, W.H., Topa, S., Xu, X.X., Zeng, X.W., Zhao, Z.X., Wang, M.K., Chen, W., Wang, F., Cheng, Y.B., He S.H. 2014, ACS Applied Materials & Interfaces, p. 10614.
20. Machine learning for molecular and materials science. Butler, K..T, Davies, D.W., Cartwright, H., Isayev, O., Walsh, A. 2018, Nature, p. 547.
21. Highly efficient photocathodes for dye-sensitized tandem solar cells. Nattestad, A., Mozer, A.J., Fischer M.K., Cheng Y.B., Mishra A., Bäuerle P., Bach U. 2010, Nature Materials, p. 31.
22. Swart, M., Bickelhaupt, M. Duran, M. Excitation energies, DFT Popularity Poll . s.l. : http://www.marcelswart.eu/dft-poll/, 2018. web link .
23. Coumarin dyes for dye-sensitized solar cells: A long-range-corrected density functional study. Wong, B.M., Cordaro, J. G. 2008, The Journal of Chemical Physics, p. 214703.
24. Evaluating the Performance of DFT Functionals in Assessing the Interaction Energy and Ground-State Charge Transfer of Donor/Acceptor Complexes: Tetrathiafulvalene−Tetracyanoquinodimethane (TTF−TCNQ) as a Model Case. Sini, G., Sears, J. S. and Brédas, J.-L.,. 2011, Journal of Chemical Theory and Computation, p. 602.