



## ZADÁNÍ BAKALÁ SKÉ PRÁCE

<b>Název:</b>	Distribuovaná infrastruktura pro virtuální screening molekul
<b>Student:</b>	Št pán Srše
<b>Vedoucí:</b>	Mgr. Jan Starý, Ph.D.
<b>Studijní program:</b>	Informatika
<b>Studijní obor:</b>	Informa ní technologie
<b>Katedra:</b>	Katedra po íta ových systém
<b>Platnost zadání:</b>	Do konce zimního semestru 2017/18

### Pokyny pro vypracování

1. Nastudujte formáty pro popis virtuálních molekul a velkých soubor chemických dat.
2. Seznamte se s distribuovanou výpo etní infrastrukturou pro výpo ty chemických vlastností (ORCA, voliteln další).
3. Vytvo te nástroj pro distribuované výpo ty nad touto infrastrukturou, který umožní vyhodnocovat výsledky takových výpo t .
4. Vytvo te nástroje pro monitorování a ov te škálovatelnost použité infrastruktury.
5. Vytvo te nástroje pro p enos, ukládání a zálohování použitých a vypo tených dat.
6. Dbejte na korektnost, ítelnost, rozši ítelnost a p enositelnost kódu.
7. Vytvo te uživatelskou dokumentaci ve form standardní manuálové stránky.

### Seznam odborné literatury

Dodá vedoucí práce.

L.S.

prof. Ing. Róbert Lórencz, CSc.  
vedoucí katedry

prof. Ing. Pavel Tvrdík, CSc.  
d kan

V Praze dne 30. zá í 2016



ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE  
FAKULTA INFORMAČNÍCH TECHNOLOGIÍ  
KATEDRA POČÍTAČOVÝCH SYSTÉMŮ



Bakalářská práce

## **Distribuovaná infrastruktura pro virtuální screening molekul**

*Bc. Štěpán Sršeň*

Vedoucí práce: Mgr. Jan Starý, Ph.D.

10. ledna 2017



---

## Poděkování

Za předávání praktických zkušeností, užitečné rady, osobní přístup a věnovaný čas děkuji vedoucímu této práce Mgr. Janu Starému, Ph.D.



---

## Prohlášení

Prohlašuji, že jsem předloženou práci vypracoval(a) samostatně a že jsem uvedl(a) veškeré použité informační zdroje v souladu s Metodickým pokynem o etické přípravě vysokoškolských závěrečných prací.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona, ve znění pozdějších předpisů. V souladu s ust. § 46 odst. 6 tohoto zákona tímto uděluji nevýhradní oprávnění (licenci) k užití této mojí práce, a to včetně všech počítačových programů, jež jsou její součástí či přílohou, a veškeré jejich dokumentace (dále souhrnně jen „Dílo“), a to všem osobám, které si přejí Dílo užít. Tyto osoby jsou oprávněny Dílo užít jakýmkoli způsobem, který nesnižuje hodnotu Díla, a za jakýmkoli účelem (včetně užití k výdělečným účelům). Toto oprávnění je časově, teritoriálně i množstevně neomezené. Každá osoba, která využije výše uvedenou licenci, se však zavazuje udělit ke každému dílu, které vznikne (byť jen zčásti) na základě Díla, úpravou Díla, spojením Díla s jiným dílem, zařazením Díla do díla souborného či zpracováním Díla (včetně překladu), licenci alespoň ve výše uvedeném rozsahu a zároveň zpřístupnit zdrojový kód takového díla alespoň srovnatelným způsobem a ve srovnatelném rozsahu, jako je zpřístupněn zdrojový kód Díla.

V Praze dne 10. ledna 2017

.....

České vysoké učení technické v Praze

Fakulta informačních technologií

© 2017 Štěpán Sršeň. Všechna práva vyhrazena.

*Tato práce vznikla jako školní dílo na Českém vysokém učení technickém v Praze, Fakultě informačních technologií. Práce je chráněna právními předpisy a mezinárodními úmluvami o právu autorském a právech souvisejících s právem autorským. K jejímu užití, s výjimkou bezúplatných zákonných licencí, je nezbytný souhlas autora.*

### **Odkaz na tuto práci**

Sršeň, Štěpán. *Distribuovaná infrastruktura pro virtuální screening molekul*. Bakalářská práce. Praha: České vysoké učení technické v Praze, Fakulta informačních technologií, 2017.

Dostupný z WWW: (<https://srsenste@bitbucket.org/srsenste/vms.git>).



---

## Abstrakt

Tato práce se zabývá distribuovanými výpočty pro virtuální screening molekul. Cílem je vytvoření software, který bude paralelně operovat nad kvantovým výpočetním programem a bude filtrovat molekuly podle jejich vlastností. Program musí být přenositelný mezi unixovými systémy, přičemž cíleným hardware jsou výpočetní klastry. První část práce je věnována rešerši souvisejících témat, na jejichž základě jsou poté v další části voleny vhodné technologie pro implementaci programu. Dále následuje samotná implementace a v závěru je program podroben extenzivnímu testování.

**Klíčová slova** absorpce světla, fluorescence, intenzita přechodu, molekula, ORCA, paralelní výpočty, UNIX, virtuální screening, výpočetní klastr

---

## Abstract

This work is focused on parallel computing for virtual molecular screening. The main goal is to develop software which will cooperate with quantum computational chemistry software and filter molecules with desired material properties. The program has to be portable among Unix-like operating systems. In the first part of this work a literature review is conducted. In the

second part acquired knowledge is used to design the software which is then extensively tested at the end of this work.

**Keywords** computer cluster, fluorescence, molecule, ORCA, parallel computing, light absorption, transition intensity, UNIX, virtual screening

---

# Obsah

Úvod	1
<b>1 Cíl práce</b>	<b>3</b>
<b>2 Virtuální screening pro výzkum materiálů</b>	<b>5</b>
2.1 Světlo ve fyzice molekul . . . . .	6
2.2 Solární organické články . . . . .	9
2.3 Organické elektroluminiscenční diody . . . . .	11
<b>3 Úvod do kvantové mechaniky molekul</b>	<b>15</b>
3.1 <i>Ab initio</i> metody . . . . .	17
3.2 Semiempirické metody . . . . .	20
3.3 DFT metody . . . . .	21
<b>4 Analýza současných řešení</b>	<b>23</b>
4.1 Formáty pro ukládání struktur molekul . . . . .	23
4.2 Kvantové výpočetní programy . . . . .	27
4.3 Zdroje dat . . . . .	29
4.4 Existující implementace virtuálního screeningu . . . . .	30
<b>5 Návrh</b>	<b>31</b>
5.1 Uživatelský vstup programu . . . . .	33
5.2 Použití programu . . . . .	36
<b>6 Implementace</b>	<b>39</b>
6.1 ORT . . . . .	39
6.2 ORI . . . . .	40
6.3 VMS . . . . .	42
<b>7 Testování a monitoring</b>	<b>45</b>

<b>Závěr</b>	<b>53</b>
<b>Literatura</b>	<b>55</b>
<b>A Seznam použitých zkratk</b>	<b>61</b>
<b>B Obsah přiloženého CD</b>	<b>63</b>

---

## Seznam obrázků

2.1	Schematické znázornění iteračního použití virtuálního screeningu. [1]	5
2.2	Spektrum elektromagnetického záření s vyznačením jednotlivých oblastí. [2]	7
2.3	Jabloňského diagram znázorňující přechody v molekule mezi singletovými stavy $S_0$ , $S_1$ a $S_2$ a tripletovým stavem $T_1$ . [3]	8
2.4	Schematické znázornění Franckova–Condonova principu na zářivých přechodech absorpce a fluorescence. Dochází k vertikálnímu přechodu do vyšší vibrační hladiny a nikoliv do energetického minima daného elektronového stavu.	9
2.5	Nejlepší ověřené účinnosti přeměny energie organických fotovoltaických článků v období od roku 2000 do současnosti. [4]	10
2.6	Základní schéma organického fotovoltaického článku. Světlo je absorbováno donorem (červená), vytvořený exciton difunduje na rozhraní s akceptorem (modrá), kde elektron přejde do vodivostního pásu akceptoru a dále do obvodu. [5]	11
2.7	Schematické znázornění procesů v OLED. Dioda je neprve dopována elektrony a elektronovými dírami, které jsou transportovány organickými vrstvami až do emitující vrstvy. Tam se nosiče náboje párují za vzniku excitonu, jehož následným zánikem dochází k vyzařování světla. [6]	12
3.1	Ukázka lokálního minima (vlevo) a sedlového bodu (vpravo) na hyperploše potenciální energie.	17
4.1	3D struktura molekuly methanu. [7]	23
5.1	Schéma rozdělení VMS programu na dílčí podprogramy.	32
5.2	Schematické znázornění využití VMS pro selekci molekul se žádanými fotochemickými vlastnostmi.	37

7.1	Závislost výpočetního času a režie VMS na počtu jader pro DFT výpočet excitovaných stavů s 1024 molekulami. . . . .	47
7.2	Závislost výpočetního času a režie VMS na počtu jader pro optimalizaci geometrií semiempirickým výpočtem s 4096 molekulami. .	47
7.3	Závislost výpočetního času a režie VMS na počtu molekul pro DFT výpočet excitovaných stavů s 12 jádry. . . . .	48
7.4	Závislost výpočetního času a režie VMS na počtu molekul pro optimalizaci geometrií semiempirickým výpočtem s 12 jádry. . . . .	49
7.5	Závislost výpočetního času a režie VMS na počtu molekul pro optimalizaci geometrií semiempirickým výpočtem s 12 jádry. Srovnání s výpočtem bez -w přepínače (index 2). . . . .	49
7.6	Závislost DFT výpočtů excitovaných stavů na počtu kontrahovaných Gaussových funkcí v bázi pro 20 molekul a skupiny bází CC a DEF2. . . . .	50
7.7	Závislost DFT výpočtů excitovaných stavů a semiempirických optimalizačních výpočtů na uvedené paměti. Ukázka na jedné molekule.	50
7.8	Závislost výpočtů excitovaných stavů na počtu zahrnutých excitovaných stavů pro DFT a semiempirickou metodu s 20 molekulami.	51
7.9	Závislost výpočtů excitovaných stavů na parametru <code>maxdim</code> pro DFT a semiempirickou metodu s 20 molekulami. . . . .	51

---

# Úvod

Rychlost vývoje vědy obecně roste exponenciální závislostí, což platí i pro odvětví chemie, kde již v současné době není možné při hledání sloučenin s požadovanými vlastnostmi experimentálně otestovat všechny potenciální kandidáty. Ukazuje se, že v mnoha případech je žádoucí celý proces automatizovat, ať už se jedná o hledání materiálů pro fotovoltaické články či účinných látek ve farmaceutickém průmyslu.

Jednou z možností je robotizace celé procedury, kdy se jedná o tzv. testování s vysokou propustností označované zkratkou HTS (high-throughput screening). Alternativou a zároveň doplňkem této metody je testování látek *in silico*, kdy se jedná o tzv. virtuální screening (VS). Podstatou VS je výpočet vlastností potenciálně žádoucích sloučenin a redukce jejich počtu pro vlastní experiment, přičemž je běžné snížení ze statisíců až na stovky molekul. Zatímco HTS má podstatně vyšší náklady, hlavně pořizovací, aplikovatelnost VS roste zároveň s rozvojem výpočetní techniky. Vzhledem k množství počítaných molekul je však standardně nutné použití výpočetně méně náročných metod, což vede k určitému procentu špatně posouzených molekul. Dalším problémem je získání prvotní databáze zkoumaných molekul, které by mohly mít požadované vlastnosti. Existuje řada přístupů od testování komerčních databází až po sofistikované metody zahrnující evoluční algoritmy. Vhodnost jednotlivých řešení je velmi úzce spjata s konkrétní aplikací VS.

V současné době je VS používán převážně ve farmaceutickém průmyslu a to takovým způsobem, že samotný pojem virtuálního screeningu bývá často zaměňován s touto jeho konkrétní aplikací. To je dáno masivním financováním farmaceutickými firmami, které tak nechávají jen málo prostoru pro akademický výzkum v této podoblasti. VS lze však v principu použít v jakémkoliv odvětví chemie, což dává možnost vzniku této bakalářské práce, která se zabývá využitím VS pro získání sady molekul s požadovanými fotochemickými vlastnostmi. Práce vznikla na základě požadavků akademické obce z oblasti výpočetní chemie, konkrétně pak díky konzultacím s prof. RNDr. et Bc. Petrem Slavíčkem, Ph.D.





---

## Cíl práce

Software vytvořený v rámci této práce by měl na základě vstupních požadavků vzít příslušnou část databáze látek, spočítat jejich zkoumané fotochemické vlastnosti a vrátit ty sloučeniny, které budou splňovat zadané podmínky. Program by měl také umět zoptimalizovat geometrie molekul uložených v databázi do jejich energetického minima. Jednoduše by tak mělo být možné získat zadaný počet molekul nejlépe splňujících definovanou intenzitou absorpce či emise elektromagnetického záření v zadané oblasti vlnových délek. Samotné chemické výpočty budou prováděny pomocí kooperujícího kvantově výpočetního programu. Software by měl nalézt uplatnění ve výzkumu fotoaktivních materiálů a měl by být rozšiřitelný na zkoumání dalších materiálových vlastností. Hlavní cílovou skupinou programu je akademická obec, zejména pak v oblasti spektroskopie a materiálového inženýrství.

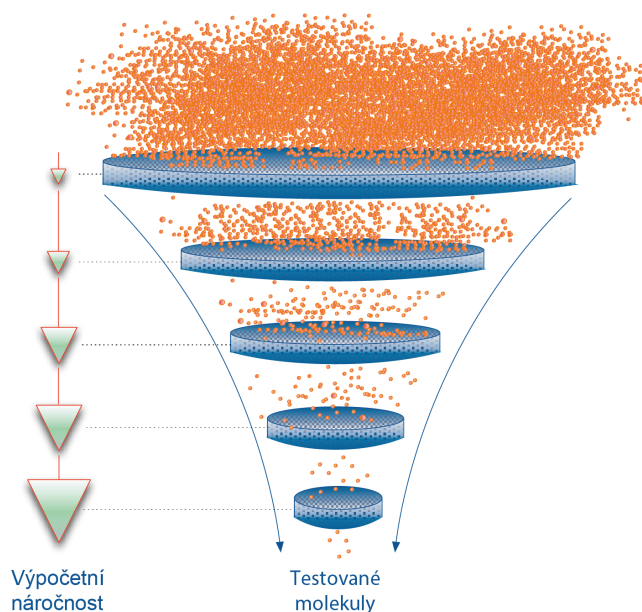
Nutnou vlastností tohoto programu je jeho přenositelnost mezi unixovými distribucemi a také jeho efektivní paralelizace. Jedná se totiž o velmi časově náročné výpočty, které jsou typicky hardwarově cílené na výpočetní klastry. Vzhledem k této skutečnosti je také nutné řádné testování škálování výpočetního času s různými parametry výpočtů. V neposlední řadě musí být možné použít výstup z jednoho běhu programu jako vstup pro další výpočty.

Na tuto práci bude navazovat bakalářská práce kolegy Jonatana Matějky, který by měl implementovat genetické algoritmy pro generování vstupních databází molekul a naprogramovat grafické prostředí programu.



## Virtuální screening pro výzkum materiálů

Jak bylo popsáno v úvodu, VS neslouží k získání jedné ideální molekuly pro dané použití. VS je koncept síta, který používá relativně nepřesné, ale zato levné metody. Je tak nevyhnutelné, že některé molekuly budou falešně pozitivní či negativní, a vytříděné molekuly je proto třeba dále zkoumat. K tomu mohou sloužit jak pokročilejší výpočetní metody, tak experiment. VS lze s výhodou používat iteračně, kdy v každém běhu zpřísníme podmínky, ale zpřesníme metodu (viz obrázek 2.1). [1]



Obrázek 2.1: Schematické znázornění iteračního použití virtuálního screeningu. [1]

Zde vyvinutý software je zaměřen na fotochemické vlastnosti a v principu by mohl najít uplatnění ve všech výrobních a výzkumných procesech zahrnujících absorpci či emisi záření, tak jak jsou popsány v následující sekci 2.1.

Jedním z takovýchto hlavních výzkumů je aktuálně například vývoj organických fotovoltaických článků. Světová spotřeba elektrické energie roste, zásoby fosilních paliv se ztenčují a při jejich spalování se uvolňují skleníkové plyny, hlavně pak oxid uhličitý. Vyřešení globálního energetického problému je jeden z nejpalčivějších problémů současnosti a organické solární články nabízejí jednu z možností, jak tento problém redukovat. [5]

Dalším příkladem uplatnění je vývoj displejů na bázi organických elektroluminiscenčních diod (OLED), které mají oproti starším displejům založeným na technologii tekutých krystalů (LCD) celou řadu výhod. OLED displeje už v tuto chvíli představují nejlepší reálně používanou technologii a vyskytují se v nejvyšších modelech telefonů, televizorů a nositelné elektroniky. OLED technologie by také mohla představovat budoucnost běžného osvětlení. [8]

## 2.1 Světlo ve fyzice molekul

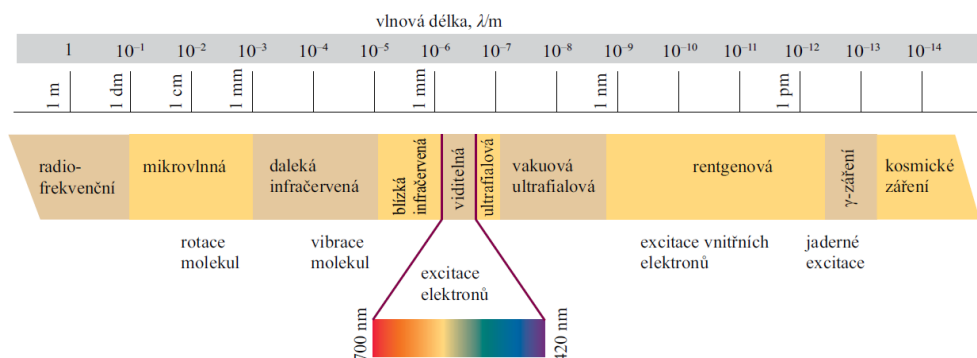
Elektromagnetické záření je molekulami pohlcováno či vyzařováno po kvantech, které nazýváme fotony. Jejich energie  $E_f$  je dána součinem Planckovy konstanty  $h$  a frekvence záření  $\nu$ . Pokud má dojít v molekule k zářivému přechodu mezi dvěma energetickými hladinami  $E_1$  a  $E_2$ , pak musí podle Bohrovy rezonanční podmínky být absorbováno/emítováno záření s energií přesně odpovídající rozdílu těchto hladin: [9]

$$E_2 - E_1 = E_f = h\nu. \quad (2.1)$$

V závislosti na energii pohlceného/vyzářeného fotonu dochází k přechodu mezi rotačními, vibračními nebo elektronovými hladinami. Elektromagnetické spektrum se zanesenými oblastmi jednotlivých přechodů je zachyceno na obrázku 2.2. Pro tuto práci jsou významné hlavně elektronové přechody odpovídající viditelnému a ultrafialovému záření.

Pro další pokračování je nutné si nejprve objasnit pojmy spin a spinová multiplicita. Spin je vnitřní moment hybnosti vyjádřený spinovým kvantovým číslem  $s$ , které je pro elektron  $\frac{1}{2}$ . V molekule se spiny nespárovaných elektronů mohou konstruktivně nebo destruktivně skládat do celkového spinu  $S$ . To se reálně projevuje v naměřených spektrech vlastností nazývanou multiplicita. Ta se spočítá jako  $2S+1$  a jedná se o stupeň degenerace elektronového kvantového stavu. Ve spektru je pak vidět rozštěpení daného stavu na  $2S+1$  velmi blízkých maxim, jelikož každý z těchto degenerovaných stavů má jinou prostorovou orientaci a tudíž jinak interaguje s celkovým orbitálním momentem hybnosti (ten je dán pohybem elektronů kolem jader) nebo ve vnějším magnetickém poli. Pokud má molekula všechny elektrony spárované, pak je její multiplicita  $2 \times 0 + 1 = 1$ , takovému stavu se říká singlet. Pokud má molekula 1 nepárový

## 2.1. Světlo ve fyzice molekul



Obrázek 2.2: Spektrum elektromagnetického záření s vyznačením jednotlivých oblastí. [2]

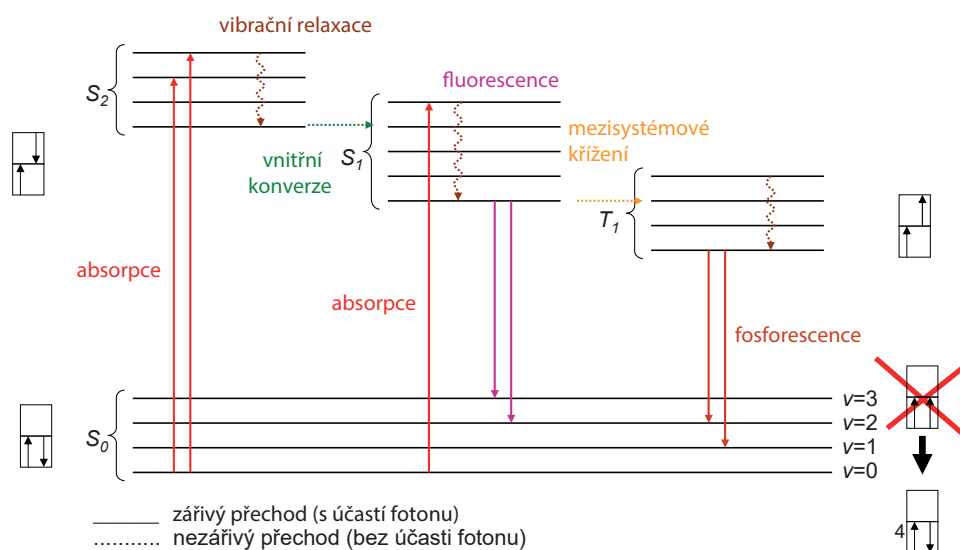
elektron, pak má multiplicitu  $2 \times \frac{1}{2} + 1 = 2$  neboli se jedná o dublet. V případě dvou nespárovaných elektronů mohou nastat dvě situace. Buďto jsou spiny orientovány stejně, pak  $S = 2 \times \frac{1}{2} = 1$  a multiplicita je 3 neboli triplet, nebo jsou orientovány opačně a vyruší se, čili se jedná opět o singlet s multiplicitou 1. V případě většího počtu nespárovaných elektronů je skládání obdobné, jen existuje více možností. [2]

Pro znázornění základních fotochemických procesů se používá Jablůnskiho diagram, jehož schéma je na obrázku 2.3. Fotochemické procesy v molekule lze rozdělit na zářivé a nezářivé podle toho, zda zahrnují foton či nikoliv. Mezi nezářivé procesy patří vibrační relaxace, což je přeměna vibrační energie na tepelnou. Dále sem patří vnitřní konverze, což je přechod mezi izoenergetickými vibračními hladinami různých elektronových stavů se stejnou multiplicitou, a mezisystémové křížení, u kterého oproti vnitřní konverzi dochází ještě ke změně multiplicity. [10]

Zářivé přechody jsou tři. Prvním z nich je absorpce, kdy je foton pohlcen a dochází k excitaci. Další dva zahrnují emisi fotonu. Jedná se o fluorescenci, kdy dochází k přechodu mezi stavy se stejnou multiplicitou, a fosforescenci, kdy dochází ke změně multiplicity. Emitované záření má standardně větší vlnové délky než záření absorbované, jelikož část energie je přeměna na tepelnou a předána okolním molekulám. [10]

Dle výběrových pravidel nesmí při přechodu dojít ke změně multiplicity. U lehkých atomů je tak mezisystémové křížení a fosforescence zakázána. U těžších jader se však začíná projevovat efekt spin-orbitální interakce, který tyto přechody umožňuje. Díky tomu tak v praxi odlišit fluorescenci od fosforescence podle doby spontánní emise. U fluorescenci se jedná o dobu maximálně v řádu mikrosekund, zatímco u fosforescence to mohou být v extrémních případech sekundy až hodiny, protože je velmi malá pravděpodobnost přechodu. [2]

Na elektronové přechody se také vztahuje tzv. Franckův–Condonův princip. Ten říká, že v okamžiku elektronového přechodu dochází pouze ke změně



Obrázek 2.3: Jabloňského diagram znázorňující přechody v molekule mezi singletovými stavy  $S_0$ ,  $S_1$  a  $S_2$  a tripletovým stavem  $T_1$ . [3]

uspořádání elektronů, nikoliv však jader. To je způsobeno velkým rozdílem v jejich hmotnostech, kdy se jádra, zjednodušené řečeno, nedokáží tak rychle přizpůsobit. Kvůli tomu dochází k tzv. vertikálnímu přechodu, jak je ukázáno na obrázku 2.4. To znamená, že jádra se po elektronovém přechodu nenacházejí v energetickém minimu daného elektronového stavu, ale zaujímají geometrii jako před excitací, ta však v kontextu nového elektronového stavu odpovídá vyšší vibrační hladině. [2]

Pro vyjádření intenzity absorpce záření na molekulární úrovni se používá oscilátorová síla  $f_{osc}$ , která udává pravděpodobnost absorpce fotonu o dané vlnové délce molekulou a pro nedegenerované stavy nabývá hodnot od 0 do 1. Oscilátorovou sílu je možné přepočítat na molární absorpční koeficient  $\epsilon$ , což je makroskopická veličina, pomocí následujícího vzorce:

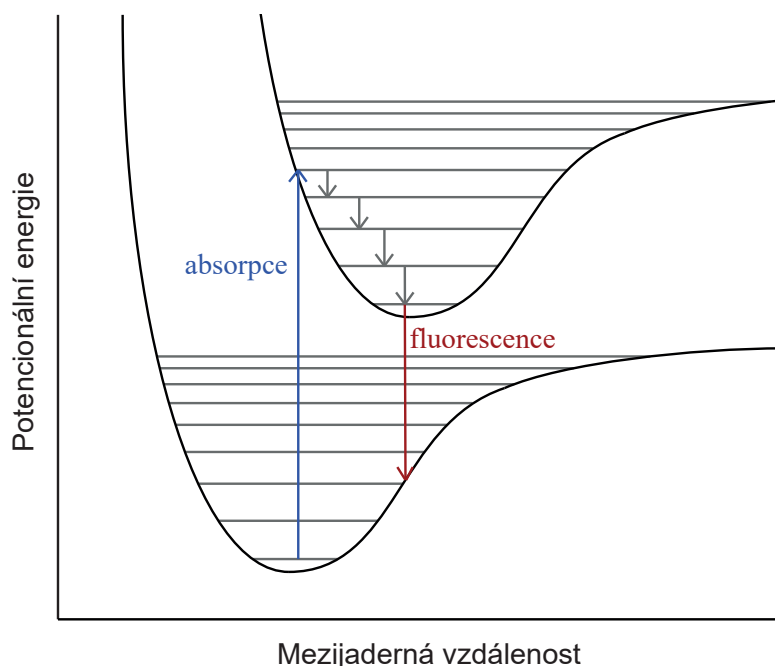
$$\epsilon = \frac{N_A \pi e^2}{2 \ln(10) \epsilon_0 m_e c}, \quad (2.2)$$

kde  $N_A$  je Avogadrova konstanta,  $e$  je elementární náboj,  $\epsilon_0$  je permitivita vakua,  $m_e$  je hmotnost elektronu a  $c$  je rychlost světla. [11]

Z molárního absorpčního koeficientu lze již jednoduše zjistit transmitanci  $T$ , čili kolik procent záření projde vzorkem o koncentraci  $c$  skrze květu o délce  $l$ :

$$T = \frac{I}{I_0} = 10^{-\epsilon c l}, \quad (2.3)$$

$I_0$  je původní intenzita záření a  $I$  je intenzita po průchodu vzorkem.



Obrázek 2.4: Schematické znázornění Franckova–Condonova principu na zářivých přechodech absorpce a fluorescence. Dochází k vertikálnímu přechodu do vyšší vibrační hladiny a nikoliv do energetického minima daného elektronového stavu.

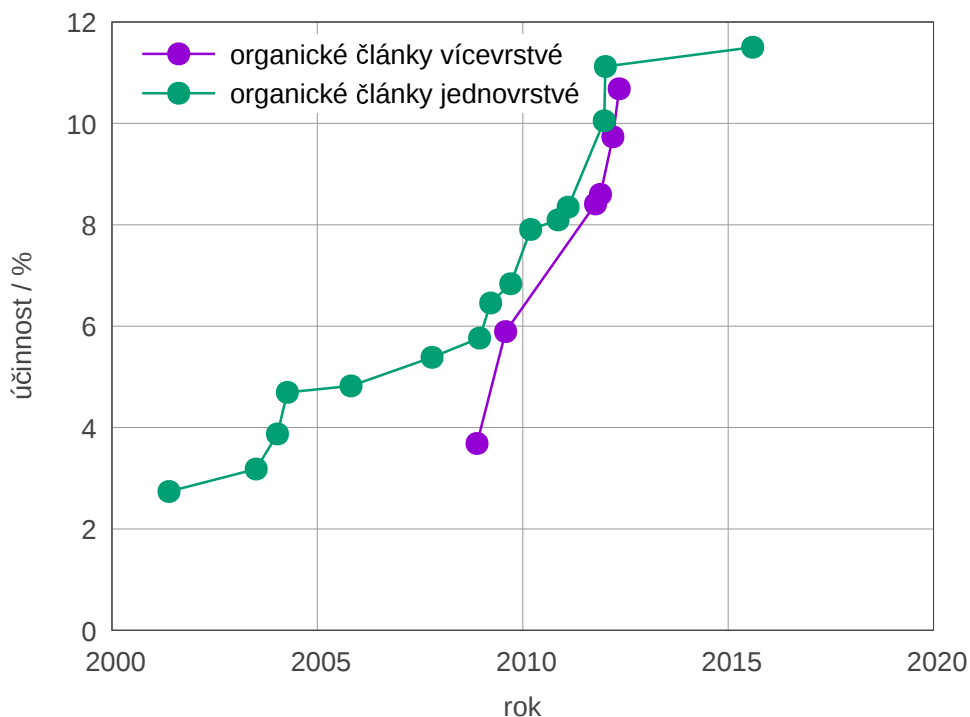
## 2.2 Solární organické články

Fotovoltaický článek je speciální polovodičová dioda, která převádí světlo na stejnosměrný proud. V závislosti na použitém materiálu mohou fotovoltaické články absorbovat kromě viditelného světla také infračervené či ultrafialové záření. [12]

Nejlepší vícepřechodové anorganické články dosahují účinnosti konverze solární energie až 44 %, používané křemíkové články dosahují až 25 % u monokrystalické formy a až 20 % u levnější polykrystalické formy. [4]. To jsou hodnoty zdaleka přesahující organické fotovoltaické články (viz obrázek 2.5), ovšem účinnost není jediné kritérium pro zhodnocení jednotlivých technologií. Jedním z nejdůležitějších faktorů pro široké nasazení technologie je její cena. [5]

Anorganické solární panely jsou limitovány dostupností materiálů, jejich toxicitou a vysokými výrobními náklady. Při jejich výrobě je používáno nákladné vakuum a kvůli nízké intenzitě absorpce světla je nutností použití velkého množství materiálu, což se projevuje nejen na ceně, ale i na hmotnosti těchto panelů. Naproti tomu organické fotovoltaické články mají nízkou pořizovací cenu a díky velké intenzitě absorpce stačí tenká vrstva materiálu, tudíž

## 2. VIRTUÁLNÍ SCREENING PRO VÝZKUM MATERIÁLŮ



Obrázek 2.5: Nejlepší ověřené účinnosti přeměny energie organických fotovoltaických článků v období od roku 2000 do současnosti. [4]

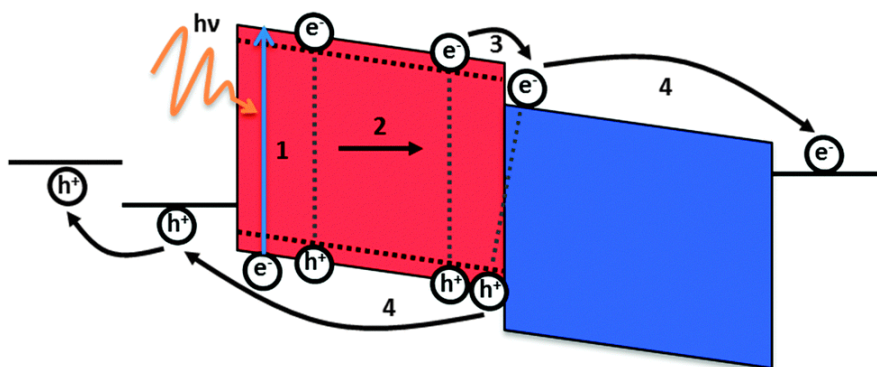
jsou tyto články lehké a flexibilní a jejich použití je tak možné na rozsáhlých plochách a pohyblivých zařízeních. [5]

Jednou ze dvou hlavních metrik pro posouzení různých technologií je doba energetické návratnosti článku [5]. Zatímco doba energetické návratnosti používaných křemíkových solárních panelů je odhadována v řádu roků [13], u organických fotovoltaických článků se jedná řádově o dny [14].

Materiál, který absorbuje v organickém článku solární energii, má typicky velký konjugovaný systém, tzn. jsou v něm střídavě jednoduché a dvojně vazby. Uhlíkové  $p_z$  orbitály tvoří delokalizovaný vazebný  $\pi$  orbital a nevazebný  $\pi^*$ , přičemž  $\pi$  orbital je zároveň nejvyšším obsazeným molekulovým orbitalem (HOMO) a  $\pi^*$  je nejnižším neobsazeným molekulovým orbitalem (LUMO). V organických solárních člancích představuje HOMO orbital valenční pás, LUMO orbital vodivostní pás a jejich energetický rozdíl šířku zakázaného pásu. [15]

Světlo s energií větší nebo stejnou, jako je šířka zakázaného pásu, může být absorbováno. Nadbytečná energie je však přeměněna na teplo, což snižuje napětí a tedy i účinnost konverze. Když je takovýmto materiálem absorbován foton, vznikne excitovaný stav, který je svázan s danou molekulou nebo částí polymerního řetězce. Tento excitovaný stav představuje exciton neboli vázaný





Obrázek 2.6: Základní schéma organického fotovoltaického článku. Světlo je absorbováno donorem (červená), vytvořený exciton difunduje na rozhraní s akceptorem (modrá), kde elektron přejde do vodivostního pásu akceptoru a dále do obvodu. [5]

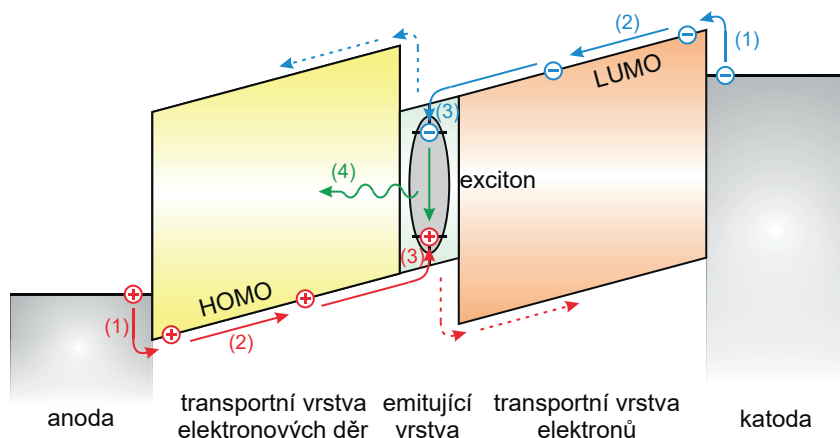
pár elektron-díra, pro vytvoření stejnosměrného proudu je nutné tento náboj uvolnit. Organické polovodiče obsahují dvousložkový systém složený z absorbojícího donoru a elektronového akceptoru, kdy je elektron uvolněn přechodem z vodivostního pásu donoru do vodivostního pásu akceptoru. Podmínkou je nižší energie vodivostního pásu akceptoru oproti vodivostnímu pásu donoru. Celý proces je schematicky znázorněn na obrázku 2.6. [15]

Moderní vícevrstvé architektury organických fotovoltaických článků obsahují více aktivních vrstev a umožňují tak absorpci fotonů o různých vlnových délkách. U dvouvrstvého článku tak má první vrstva velkou šířku zakázaného pásu a zachytává vysokoenergetické fotony, zatímco druhá vrstva absorbuje nízkenergetické fotony, které první vrstvou projdou. [16] Jedná se o relativně novou technologii a tak její účinnost zatím dosáhla pouze 12 %, ale v blízké budoucnosti se odhaduje její nárůst, přičemž by měla překonat jednovrstvé články zhruba o 30 %. [5]

Výsledná účinnost konverze organického článku je vysoce závislá na vlastnostech absorbojícího donoru a akceptoru v aktivní vrstvě. Základním parametrem vhodnosti materiálů je jejich šířka zakázaného pásu a intenzita absorpce a právě zde by měl najít uplatnění zde vyvíjený program. Většina používaných materiálů v tuto chvíli pochází z výzkumu OLED displejů, jelikož se jednalo o hlavní komunitní téma při počátku vývoje organických solárních článků. [5]

## 2.3 Organické elektroluminiscenční diody

OLED displeje mají celou řadu výhod oproti starším LCD displejům. Mají nízkou spotřebu energie, poskytují ostřejší obraz a větší pozorovací úhly, mají lepší kontrast a dokáží pracovat ve velkém teplotním rozmezí. Dále jsou velmi lehké a flexibilní, mohou být průhledné a jsou tisknutelné, díky čemuž je lze



Obrázek 2.7: Schematické znázornění procesů v OLED. Dioda je neprve dopována elektrony a elektronovými dírami, které jsou transportovány organickými vrstvami až do emitující vrstvy. Tam se nosiče náboje párují za vzniku excitonu, jehož následným zánikem dochází k vyzáření světla. [6]

použít veliké plochy a nebo na zakřivená či pohyblivá mobilní zařízení. Také mají oproti LCD přibližně o tři řády lepší odezvu, což z nich dělá ideální technologii pro 3D zobrazování. U OLED displejů existuje také možnost kontrolovat intenzitu každého pixelu zvlášť narozdíl od LCD technologie. [8]

OLED technologie je založena na elektroluminiscenčním jevu, ten byl poprvé pozorován v organických krystalech v roce 1963 [17]. Velký průlom byl poté udělán v roce 1987, kdy Tang a Van Slyke z firmy Kodak připravili účinné OLED pracující za nízkého napětí [18]. V principu může být OLED složená pouze z jedné organické vrstvy vložené mezi elektrody. V praxi se nicméně používá vícevrstvá architektura, která má větší účinnost a delší životnost, jelikož vlastnosti jednotlivých vrstev je možné ovlivňovat vhodným výběrem materiálů. [6]

Na obrázku 2.7 jsou schematicky zachyceny procesy v architektuře se třemi organickými vrstvami. Na elektrody je přivedeno externí napětí o hodnotě několika voltů a z katody tak proudí elektrony do LUMO orbitalu elektronové transportní vrstvy, zatímco anoda dopuje elektronovými dírami HOMO orbital transportní vrstvy elektronových děr. Tyto částice putují k sobě skrze organické transportní vrstvy až vytvoří v emitující vrstvě exciton neboli vázaný pár elektron-díra. Exciton může následně zaniknout za vyzáření fotonů. [6]

V dnešní době jsou, stejně jako u organických solárních článků, vyvíjeny architektury s více aktivními vrstvami [19]. Emise záření v OLED může být založena na fluorescenci i fosforescenci, přičemž fosforescence vykazuje větší účinnost, ovšem zatím je technologicky pozadu. Výsledná barva pixelu je součtem záření z jednotlivých barevných diod či emisních vrstev a skládá se z červené, zelené a modré. Současné emisní materiály pro modrou barvu vykazují

### 2.3. Organické elektroluminiscenční diody

---

výrazně horší vlastnosti než pro zbývající dvě. Mají podstatně kratší životnost, menší účinnost elektroluminiscence a kvalitu barvy. [20] Právě pro hledání vhodného elektroluminiscenčního materiálu pro modrou barvu by také mohl být použit program vyvíjený v této práci.



# Úvod do kvantové mechaniky molekul

Pro pochopení následujícího textu a správný návrh programu je nutné nejprve provést rešerši esenciálních mechanismů kvantové teorie molekul. Základem kvantové mechaniky je vlnově-částicový dualismus a kvantování fyzikálních veličin. Trajektorie částice je nahrazena vlnou, která je reprezentována tzv. vlnovou funkcí  $\Psi$ . Pokud známe vlnovou funkci molekuly, můžeme odvodit její další vlastnosti. Druhá mocnina vlnové funkce na daném intervalu je také úměrná pravděpodobnosti výskytu částice v tomto intervalu. Roku 1926 publikoval Erwin Schrödinger diferenciální rovnici druhého řádu, která slouží ke zjištění vlnové funkce libovolně složitěho systému. Pod názvem časově závislá Schrödingerova rovnice se stala ústřední rovnicí kvantové mechaniky:

$$\hat{H}\Psi = i\hbar\frac{\partial}{\partial t}\Psi, \quad (3.1)$$

kde  $\hat{H}$  je Hamiltonův operátor neboli operátor celkové energie,  $i$  je imaginární jednotka a  $\hbar$  je redukovaná Planckova konstanta. [2]

Často však nepotřebujeme zahrnout časovou závislost a stačí nám stacionární řešení, v tu chvíli se výše uvedená rovnice zjednoduší na tvar:

$$\hat{H}\Psi = E\Psi, \quad (3.2)$$

kde  $E$  je vlastní číslo Hamiltonova operátoru  $\hat{H}$ , které představuje možné hodnoty celkové energie systému. Hamiltonův operátor je součtem kinetického a potenciálního členu, pro částici o hmotnosti  $m$  má následující tvar:

$$\hat{H} = -\frac{\hbar^2}{2m}\Delta + \hat{V}, \quad (3.3)$$

kde  $\hat{V}$  je operátor potenciální energie a  $\Delta$  je Laplaceův operátor představující sumu druhých derivací podle polohových souřadnic. [2]

### 3. ÚVOD DO KVANTOVÉ MECHANIKY MOLEKUL

---

Při absenci vnějšího pole, zanedbání relativistických efektů a některých speciálních interakcí lze Hamiltonův operátor získat z kinetické energie elektronů a z jejich vzájemné elektrostatické interakce:

$$\hat{H} = - \sum_i \frac{\hbar^2 \Delta}{2m_e} - \sum_k \frac{\hbar^2 \Delta}{2m_k} + \frac{e^2}{4\pi\epsilon_0} \left( \sum_{i<j} \frac{1}{r_{ij}} + \sum_{k<l} \frac{Z_k Z_l}{r_{kl}} - \sum_i \sum_k \frac{Z_k}{r_{ik}} \right), \quad (3.4)$$

kde  $i$  a  $j$  jsou indexy elektronů a  $k$  a  $l$  jsou indexy jader,  $m_e$  je hmotnost elektronu,  $m_k$  hmotnost konkrétního atomového jádra,  $e$  elementární náboj,  $\epsilon_0$  permitivita vakua,  $r_{mn}$  je vzdálenost daných dvou částic a  $Z_k$  a  $Z_l$  jsou náboje příslušných jader. [21]

Sestavit Hamiltonův operátor a tedy i Schrödingerovu rovnici není problém ani pro velké molekuly, problém nastává u jejího řešení. Analyticky lze řešit pouze pro nejjednodušší systémy a pro větší molekuly nelze bez dostatečných aproximací řešit ani numericky. Jednou takovou základní aproximací je oddělení pohybu jader a elektronů, jelikož je mezi nimi více než tisícinásobný hmotnostní rozdíl a jádra jsou tak relativně vůči elektronům téměř stacionární. Tato aproximace se nazývá Bornova–Oppenheimerova a lze díky ní celková vlnová funkce zapsat jako součin vlnové funkce elektronové  $\Psi_{\text{el}}$  a vlnové funkce jaderné  $\Psi_{\text{jad}}$ :

$$\Psi = \Psi_{\text{el}} \Psi_{\text{jad}}. \quad (3.5)$$

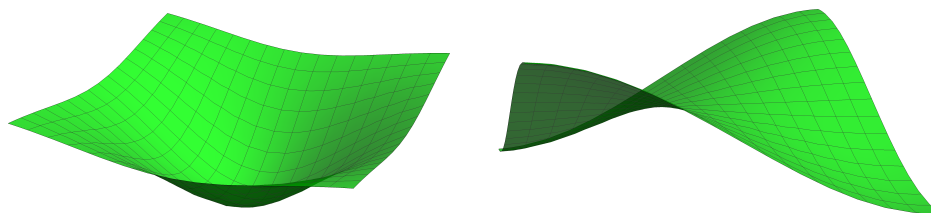
Díky tomu lze také samostatně řešit elektronová Schrödingerova rovnice pro dané geometrické uspořádání jader:

$$\hat{H}_{\text{el}} \Psi_{\text{el}} = E_{\text{el}} \Psi_{\text{el}}. \quad (3.6)$$

V elektronovém Hamiltonovu operátoru  $\hat{H}_{\text{el}}$  tak oproti rovnici 3.4 ubude druhý člen představující kinetickou energii jader a čtvrtý člen, tedy repulze jader, se stane konstatním. [22]

Nelineární molekuly mají  $3N - 6$  geometrických parametrů (lineární pak  $3N - 5$ ), kde  $N$  je počet atomů. Pro každou geometrii může být na základě Bornovy–Oppenheimerovy aproximace spočítána elektronová energie. Graf závislosti elektronové energie na vnitřních souřadnicích (vazebné vzdálenosti, vazebné úhly a dihedralní úhly) se nazývá hyperplocha potenciální energie (PES). Každý bod na PES pak představuje právě jednu geometrii molekuly a chemické transformace si tak lze představit jako pohyb po hyperploše potenciální energie. [23]

Chemicky významnými body na PES jsou lokální minima, kdy se jedná o stabilní geometrická uspořádání, a sedlové body prvního řádu, které reprezentují tranzitní stavy (viz obrázek 3.1). Optimalizace molekulové geometrie tak odpovídá hledání minima na PES, kdy v každém kroku jsou spočítány energetické gradienty původní geometrie a na jejich základě je navrhována geometrie nová. [24]



Obrázek 3.1: Ukázka lokálního minima (vlevo) a sedlového bodu (vpravo) na hyperploše potenciální energie.

Každá molekula má řadu hyperploch, jelikož se může nacházet v různých elektronově excitovaných stavech. Pro každou geometrii tak existuje spektrum elektronových energií a pomocí elektromagnetického záření o vhodné vlnové délce je pak možné mezi těmito hladinami přecházet. [21]

Na řešení Schrödingerovy rovnice se uplatňuje variační princip, podle kterého energie získaná s jakoukoliv zkusmou vlnovou funkcí bude větší nebo rovna energii reálné. Z toho plyne, že můžeme testovat rozličné funkce a ta s nižší energií popisuje lépe daný systém. Na základě variačního principu bývá vlnová funkce v rámci použité kvantové metody parametrizována a tyto parametry iteračně optimalizovány pro získání co nejnižší energie. [21]

Existuje celá řada kvantových výpočetních metod, které lze použít pro řešení Schrödingerovy rovnice, lze je však rozdělit do tří hlavních kategorií. [22]

### 3.1 *Ab initio* metody

První skupinou jsou tzv. *ab initio* (doslova „od počátku“) metody, které vychází z čistě fyzikálních zákonitostí kvantové mechaniky a nejsou v nich použity žádné empirické vztahy. *Ab initio* metody poskytují většinou nejlepší výsledky, ovšem za cenu vysoké výpočetní náročnosti. [21]

Základní *ab initio* metodou, ze které všechny ostatní vycházejí, je Hartreeho–Fockova (HF) metoda. Ta vychází z hrubé aproximace tzv. jednoelektronového přiblížení. Princip tohoto zjednodušení spočívá v nahrazení víceelektronových interakcí zprůměrovaným elektrostatickým polem. Elektrony jsou tak v HF metodě na sobě nezávislé a víceelektronová vlnová funkce se formálně rozpadá na součin jednoelektronových funkcí  $\psi_i$  také nazývaných molekulové orbitály:

$$\Psi_{\text{el}} = \prod_{i=1}^N \psi_i, \quad (3.7)$$

kde  $N$  je celkový počet elektronů v molekule. [22]

Taková vlnová funkce však nesplňuje Pauliho vylučovací princip, který říká, že nemohou existovat dva elektrony ve stejném kvantovém stavu. V praxi to znamená, že při prohození dvou elektronů ve vlnové funkci musí tato funkce

### 3. ÚVOD DO KVANTOVÉ MECHANIKY MOLEKUL

---

změnit znaménko. Je tedy nutné výše uvedenou funkci antisymetrizovat vůči permutaci elektronů, jako vhodná forma se ukázaly tzv. Slaterovy determinanty:

$$\Psi_{\text{el}} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(\vec{r}_1)\alpha(s_1) & \psi_1(\vec{r}_2)\alpha(s_2) & \dots & \psi_1(\vec{r}_N)\alpha(s_N) \\ \psi_1(\vec{r}_1)\beta(s_1) & \psi_1(\vec{r}_2)\beta(s_2) & \dots & \psi_1(\vec{r}_N)\beta(s_N) \\ \psi_2(\vec{r}_1)\alpha(s_1) & \psi_2(\vec{r}_2)\alpha(s_2) & \dots & \psi_2(\vec{r}_N)\alpha(s_N) \\ \psi_2(\vec{r}_1)\beta(s_1) & \psi_2(\vec{r}_2)\beta(s_2) & \dots & \psi_2(\vec{r}_N)\beta(s_N) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{N/2}(\vec{r}_1)\alpha(s_1) & \psi_{N/2}(\vec{r}_2)\alpha(s_2) & \dots & \psi_{N/2}(\vec{r}_N)\alpha(s_N) \\ \psi_{N/2}(\vec{r}_1)\beta(s_1) & \psi_{N/2}(\vec{r}_2)\beta(s_2) & \dots & \psi_{N/2}(\vec{r}_N)\beta(s_N) \end{vmatrix}, \quad (3.8)$$

kde  $\psi_i(\vec{r}_j)$  jsou prostorové molekulové orbitály a  $\alpha(s_j)$  a  $\beta(s_j)$  jsou ortogonální spinové funkce, které přijímají pouze dvě hodnoty spinu  $\pm 0.5$  a na základě toho vracejí hodnoty 0 nebo 1. V každém prostorovém orbitalu se tak mohou nacházet 2 elektrony s opačným spinem. Koeficient před determinan-tem normalizuje vlnovou funkci, pravděpodobnost výskytu elektronů v celém prostoru musí být totiž rovna jedné. [22]

Výše byl popsán variační princip, pro jeho použití je však nutné nejprve co nejvhodněji zvolit tvar molekulových orbitalů. Praktické se ukázalo použití lineární kombinace atomových orbitalů (LCAO)  $\phi_j$ , což jsou analyticky získané jednoelektronové vlnové funkce:

$$\psi_i = \sum_{j=1}^N a_{ij} \phi_j, \quad (3.9)$$

kde  $a_{ij}$  jsou optimalizované koeficienty. Množiny  $N$  atomových orbitalů se nazývá báze atomových orbitalů a je základním vstupním parametrem kvantových výpočetních programů. [24]

Slater původně použil sadu atomových orbitalů získaných pro atom vodíku, které jsou z fyzikálního hlediska přirozené. Nicméně pro menší náročnost matematických operací byly nahrazeny kontrahovanými Gaussovými funkcemi neboli lineárními kombinacemi Gaussových funkcí s fixními koeficienty. S počtem Gaussových funkcí v bázi roste kvalita popisu systému, avšak roste také výpočetní náročnost. Kvůli tomu bývají valenční orbitály popisovány více funkcemi než orbitály vnitřní, jelikož mají větší vliv na chemické vlastnosti. V případě, kdy jsou valenční orbitály popsány dvěma kontrahovanými Gaussovými funkcemi, jedná o tzv. valenční *double zeta* bázi. Pokud jsou popsány třemi funkcemi, jedná se o valenční *triple zeta* bázi, pro více funkcí je pak názvosloví obdobné. [25]

Použití HF a LCAO aproximací dohromady vede k jednoelektronovým Roothaanovým–Hallowým rovnicím, které mají v maticové formě následující tvar:

$$\mathbf{FC} = \mathbf{\epsilon SC}, \quad (3.10)$$



kde  $\mathbf{F}$  je Fockova matice analogická k Hamiltonovu operátoru v Schrödingrově rovnici,  $\mathbf{C}$  je matice rozvojových koeficientů molekulových orbitalů,  $\epsilon$  je diagonální matice orbitálních energií, ze kterých lze spočítat celková elektronová energie, a  $\mathbf{S}$  je překryvová matice molekulových orbitalů. Fockova matice obsahuje člen kinetické energie elektronů, elektrostatické přitahování jader a elektronů a HF potenciál, který zahrnuje dříve zmíněné repulsivní působení zprůměrovaného pole ostatních elektronů. [22]

Dosavadními úpravami v této podkapitole jsme tak převedli jednu komplexní diferenciální rovnici pro  $N$  elektronů na  $N$  relativně jednoduchých jednoelektronových lineárních rovnic, čímž se řádově snižuje výpočetní náročnost.

Fockova matice je funkcí rozvojových koeficientů, její tvar tudíž není předem znám a rovnici je třeba řešit iterativně. K tomuto účelu slouží metoda self-konzistentního pole (SCF), v praxi se jedná o ekvivalentní označení HF metody. V SCF metodě řešení Roothaanových–Hallowých rovnic jsou nejprve zvoleny počáteční rozvojové koeficienty, poté je sestrojena Fockova a překryvová matice a následně jsou řešením soustavy získány nové koeficienty a celková energie. Vypočtené koeficienty jsou znovu dosazeny do rovnice a proces se opakuje, dokud celková elektronová energie dostatečně nezkonverguje. [26]

Přesnost řešení poskytovaného HF metodou roste s velikostí báze atomových orbitalů, i při použití nekonečně velké báze se však kvůli použitým aproximacím nedostaneme pod hodnotu, která je označována jako HF limita. Rozdíl mezi HF limitou a exaktní nerelativistickou energií se nazývá korelační energie a je cílem post-HF metod tuto dodatečnou energii co nejvíce zahrnout. [21]

Prvním z přístupů je model konfigurační interakce, kdy není vlnová funkce tvořena pouze jedním Slaterovým determinantem základního stavu, ale zahrnuje i determinanty pro excitované stavy. Při zahrnutí všech možných elektronových stavů se jedná o tzv. plnou konfigurační interakci, která má následující tvar:

$$\Psi_{\text{el}} = \sum_i c_i \Psi_i, \quad (3.11)$$

kde  $i$  probíhá přes všechny elektronové stavy a  $\Psi_0$  odpovídá HF vlnové funkci základního stavu, koeficienty  $c_i$  jsou pak získány na základě variačního principu. Při použití nekonečné báze poskytuje model plné konfigurační interakce exaktní nerelativistické řešení, v praxi je však nutné omezit jak velikost báze, tak počet zahrnutých excitovaných stavů. Metody založené na modelu konfigurační interakce poskytují i energie excitovaných stavů, avšak při neúplné konfigurační interakci jsou zatíženy chybou známou jako *size consistency error*, kdy se energie dvou nekonečně vzdálených subsystémů nerovná energii těchto subsystémů izolovaných. [26]

Vylepšením modelu konfigurační interakce je teorie spřažených klastrů, kdy je celková vlnová funkce také kombinací determinantů základního a excitova-

ných stavů, ovšem v jiné matematické formě:

$$\Psi_{\text{el}} = e^{\hat{T}} \Psi_0, \quad (3.12)$$

kde  $\hat{T}$  je klastrový operátor, který z determinantu HF vlnové funkce vytvoří determinanty excitovaných stavů. Při omezení počtu zahrnutých excitovaných stavů dochází díky exponenciální formě i k nepřímému zahrnutí vyšších vícenásobných excitovaných stavů a metody založené na teorii spřažených klastrů tak netrpí na *size consistency error*. Tyto metody jsou velmi přesné a v současné době jsou považovány za standard kvality kvantové chemie. Na druhou stranu se jedná o velmi výpočetně náročné metody, přičemž náročnost nepoužívanější CCSD metody roste s šestou mocninou velikosti systému. [21]

Jiný přístup zahrnutí korelační energie využívají metody založené na poruchové teorii. Pokud známe řešení operátorové rovnice jednoduššího operátoru, lze za předpokladu malého rozdílu odhadnout vlastní funkce a vlastní čísla operátoru složitějšího. Použití tohoto principu umožňuje právě malý poměr korelační a celkové elektronové energie. Celkový Hamiltonův operátor tak lze vyjádřit jako součet nekorelovaného Hamiltonova operátoru  $\hat{H}_0$  se známým řešením a poruchy  $\hat{V}$  vynásobené bezrozměrným parametrem  $\lambda$ :

$$\left(\hat{H}_0 + \lambda\hat{V}\right) \Psi = E\Psi. \quad (3.13)$$

V rámci poruchového počtu pak dostáváme vyjádření energie i vlnové funkce ve formě nekonečných řad:

$$E = \sum_{i=0}^{\infty} \lambda^i E^{(i)} \quad (3.14)$$

$$\Psi = \sum_{i=0}^{\infty} \lambda^i \Psi^{(i)}. \quad (3.15)$$

Pokud budeme brát v potaz pouze omezený počet členů a vyřešíme soustavu rovnic, obdržíme korekci k neporušenému Hamiltonovu operátoru, jejíž řád udává maximální hodnota  $i$ . [21]

## 3.2 Semiempirické metody

Další kategorií jsou semiempirické metody, které také vycházejí z HF metody, ale obsahují dodatečné parametry, které byly typicky získány pomocí náročných *ab initio* metod nebo z experimentálních hodnot. V semiempirických metodách bývají zanedbány některé členy Hamiltonova operátoru, standardně jsou brány v potaz pouze valenční elektrony a příspěvek ostatních je nahrazen konstantami. Bázový set je zredukován na minimální valenční reprezentaci. [24]

Hlavní zjednodušení, které stojí za redukcí výpočetního času, je aproximace nulového diferenciálního překryvu, kdy jsou zanedbány některé integrály představující dvouelektronovou repulzi. Jednotlivé semiempirické metody se liší hlavně v tom, které překryvové integrály zanedbávají. Standardní je zanedbání překryvových integrálů atomových orbitalů, které přísluší různým atomům, kdy dochází k redukcí počtu integrálů vzájemné elektronové interakce v Roothaanových–Hallowých rovnicích o dva řády. Zbývající dvouelektronové integrály jsou parametrizovány, čímž je částečně vykompenzována tato aproximace. [23]

Hlavní výhodou semiempirických metod je jejich velmi nízká výpočetní náročnost. Nevýhodou je, že tyto metody sice mohou dobře fungovat pro skupiny látek, pro které byly parametrizovány, ale pro jiné mohou naprosto selhávat. [24]

### 3.3 DFT metody

Na rozmezí dvou výše uvedených kategorií stojí metody založené na teorii funkcionálu hustoty (DFT). Jejich odvození je sice čistě fyzikální a splňuje podmínky *ab initio* metod, ovšem v jejich reálné aplikaci bývají částečně parametrizovány.

DFT metody nepoužívají pro popis systému vlnovou funkci, ale elektronovou hustotu. energii základního stavu je pak možné rozepsat jako sumu kinetické energie  $E_T$ , interakční energie elektronové hustoty s jádry  $E_V$ , repulzivní energie dané distribucí elektronové hustoty  $E_J$  a výměnně-korelačního potenciálu  $E_{XC}$ :

$$E = E_T + E_V + E_J + E_{XC} \quad (3.16)$$

Každý člen na pravé straně rovnice je závislý na elektronové hustotě, celková energie je tak jejím funkcionálem. Vyjma výměnně-korelačního potenciálu umíme všechny členy vyjádřit analyticky, jeho formu však neznáme a používají se tak různé aproximace. [24]

Integrací elektronové hustoty  $\rho(\vec{r})$  přes celý prostor musíme dostat počet elektronů  $N$  v systému:

$$N = \int \rho(\vec{r}) d\vec{r}, \quad (3.17)$$

přičemž elektronovou hustotu formulujeme pomocí obsazených tzv. Kohnových–Shamových orbitalů:

$$\rho(\vec{r}) = \sum_{i=1}^N \psi_i^*(\vec{r}) \psi_i(\vec{r}). \quad (3.18)$$

Tyto orbitály jsou získány na základě Kohnových–Shamových rovnic, které mají formálně stejný tvar jako Roothaanovy–Hallowy rovnice u HF metody:

$$h_i^{KS} \psi_i(\vec{r}) = \epsilon \psi_i(\vec{r}), \quad (3.19)$$

kde  $h_i^{KS}$  je Kohnův–Shamův jednoelektronový operátor. Obdobným způsobem za pomoci SCF jsou tyto rovnice i řešeny. Nejprve odhadneme počáteční elektronovou hustotu, vyřešíme Kohnovy–Shamovy rovnice a získáme sadu orbitalů. S touto sadou spočteme novou elektronovou hustotu a postup opakujeme dokud nejsou výsledky konstantní v rámci konvergenčního kritéria. [21]

Za předpokladu znalosti podoby výměnně-korelačního funkcionálu by DFT metoda poskytovala přesnou nerelativistickou energii. V praxi jsou však používány aproximativní funkcionály, které vnášejí do DFT onu empiriku, a právě a pouze jimi se liší jednotlivé DFT metody. Tyto metody jsou podstatně výpočetně náročnější než semiempirické metody, ale stále výrazně méně náročné oproti post-HF metodám. V současné době se jedná o nejpopulárnější kvantovou výpočetní metodu, jelikož současná úroveň výpočetní techniky umožňuje jejich aplikaci i na velmi rozsáhlé systémy, přičemž je dosaženo velmi dobrého poměru kvality výsledků ku výpočetnímu času. V rámci samotné DFT metody nejsou počítány excitované stavy, k tomu slouží tzv. časově závislá DFT (TDDFT), kdy je zkoumána odezva systému na časově závislý potenciál jako je vnější elektrické pole. [27]

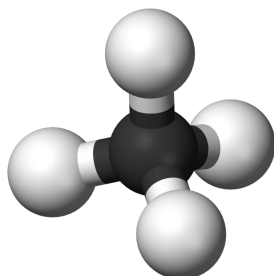
## Analýza současných řešení

### 4.1 Formáty pro ukládání struktur molekul

Počet formátů pro ukládání molekul se nevejde do jednoho sta a pokusy o jejich unifikaci zatím více či méně selhávají. Každé odvětví chemie má své specifické formáty, stejně jako má své specifické požadavky. V některých oblastech se používají pouze 2D struktury, zatímco jinde se zase nezapisují atomy vodíku a počítá se s nimi implicitně. Často také bývají struktury zapisovány jenom symbolicky pomocí vazeb, ale nejsou zachyceny reálné vzdálenosti atomů. Situaci ještě komplikují dodavatelé chemického softwaru, kteří si velmi často zavádějí každý své vlastní formáty, přesto že již o ně není v dané oblasti nouze. [28]

Pokud se však omezíme pouze na univerzální formáty používané v kvantových výpočetních programech, tak se nám toto ohromné množství zredukuje na 2 formáty. Vnitřní struktura těchto formátů je ukázána na molekule methanu, jehož 3D zobrazení je pro srovnání na obrázku 4.1.

**XYZ** Tento formát má velmi jednoduchou strukturu, kdy na prvním řádku je standardně uveden počet atomů v molekule, druhý řádek je vyhrazen pro



Obrázek 4.1: 3D struktura molekuly methanu. [7]

#### 4. ANALÝZA SOUČASNÝCH ŘEŠENÍ

---

komentář a poté již následuje výčet atomů s kartézskými souřadnicemi. První dva řádky mohou být však někdy vynechány. [29]

```
5
methan
C      0.034584    -0.177548    0.096655
H     -0.085081     0.445186    0.982823
H      1.037221    -0.044624   -0.309878
H     -0.703211     0.111147   -0.652091
H     -0.112391    -1.223326    0.366491
```

**Z-matice** Spíše než o jediný formát se jedná o způsob zápisu struktury molekul pomocí vnitřních souřadnic, tedy pomocí vazebných délek a vazebných a dihedralních úhlů (pro 4 atomy zapsané jako ABCD se jedná o úhel mezi rovinami ABC a BCD). Konkrétní zápis se však liší mezi programy. Nejstandardnější je zápis, kde první atom je jednoduše zapsán pouze svou zkratkou, druhý atom už je nadefinován spojením s prvním atomem a délkou této vazby. U třetího atomu se definuje opět pořadí atomu, ke kterému je navázán, a délka vazby, navíc se ale ještě uvádí pořadí dalšího již zapsaného atomu a úhel mezi těmito třemi atomy. Od čtvrtého atomu dále se uvádí ještě pořadí jednoho atomu a dihedralní úhel mezi danými čtyřmi atomy. [29]

```
C
H 1 1.089983
H 1 1.089983 2 109.401
H 1 1.089983 2 109.401 3 120.000
H 1 1.089983 2 109.401 3 240.000
```

U symetrických molekul, jako je zde uvedený methan, se některé konstanty opakují, proto je ve většině programech možné danou hodnotu označit textovým řetězcem, kterému se přiřadí konkrétní číslo pod výčtem atomů. [29]

```
C
H 1 vazba1
H 1 vazba1 2 uhel1
H 1 vazba1 2 uhel1 3 dihed1
H 1 vazba1 2 uhel1 3 dihed2
```

```
vazba1 1.089983
uhel1 109.401
dihed1 120.000
dihed2 240.000
```

Další formáty, které jsou pro tuto práci relevantní, jsou univerzální formáty nejvíce používané pro ukládání kompletních 3D struktur v chemických databázích. [30,31]

**CML** Jedná se o zkratku pro chemický značkovací jazyk. Tento formát je založen na rozšiřitelném značkovacím jazyku (XML) a postihuje naprostou většinu oblastí chemie. CML schéma je tedy velmi komplexní a zde bude uveden pouze způsob pro zápis 3D struktury. Po XML definici následuje prvek `molecule`, ve kterém se nachází atribut `id` s názvem molekuly a definice schématu. Element `molecule` dále obsahuje elementy `atomArray` a `bondArray`. Element `atomArray` definuje jednotlivé atomy pomocí elementu `atom`, který obsahuje atributy pro id atomu, typ prvku a kartézské souřadnice. Element `bondArray` obsahuje elementy `bond`, které definují vazby pomocí atributů pro definici vázaných atomů a řád vazby. Soubory s tímto formátem mívají příponu `.cml`. [32]

```
<?xml version="1.0"?>
<molecule id="methan"
  ↪ xmlns="http://www.xml-cml.org/schema">
<atomArray>
  <atom id="a1" elementType="C" x3="0.034584"
    ↪ y3="-0.177548" z3="0.096655"/>
  <atom id="a2" elementType="H" x3="-0.085081"
    ↪ y3="0.445186" z3="0.982823"/>
  <atom id="a3" elementType="H" x3="1.037221"
    ↪ y3="-0.044624" z3="-0.309878"/>
  <atom id="a4" elementType="H" x3="-0.703211"
    ↪ y3="0.111147" z3="-0.652091"/>
  <atom id="a5" elementType="H" x3="-0.112391"
    ↪ y3="-1.223326" z3="0.366491"/>
</atomArray>
<bondArray>
  <bond atomRefs2="a4 a1" order="1"/>
  <bond atomRefs2="a3 a1" order="1"/>
  <bond atomRefs2="a1 a5" order="1"/>
  <bond atomRefs2="a1 a2" order="1"/>
</bondArray>
</molecule>
```

**MOL** Jedná se obecně používaný formát, který v sobě zahrnuje kartézské souřadnice. Soubory mají koncovku `.mol` a jsou určeny vždy pro uložení pouze jedné molekuly. Nejprve je uvedena třířádková popisná hlavička a poté následuje samotný zápis molekuly. V hlavičce je první řádek určen pro název molekuly, druhý řádek pro iniciály autora, název programu, datum, dimensionalitu apod. a třetí řádek je jakýkoliv komentář nebo

#### 4. ANALÝZA SOUČASNÝCH ŘEŠENÍ

---

prázdná řádka. V zápisu molekuly je nejprve tzv. řádek počtů, ten obsahuje počet atomů, počet vazeb, nějaké další méně používané vlastnosti a na konci je verze formátu. Poté jsou vypsány po řádku jednotlivé atomy, kdy první tři čísla udávají kartézské souřadnice, čtvrtá je značka prvku, poté následují čísla rozdílů hmotnosti a náboje od standardního stavu a další méně podstatná čísla. Následně se zapisuje vazebná sekce, kde první dvě čísla jsou indexy atomů a třetí číslo je typ vazby, opět je možné použít i některé další vlastnosti. Na konci je sekce dalších strukturních vlastností, kde každý řádek začíná písmenem „M“, dvěma mezerami a třípísmennou zkratkou vlastnosti, zbytek řádku je zápis dané vlastnosti. Blok vlastností je zakončen vlastností „END“. [33]

```
methan
SSVMS-----12151622003D

5 4 0 0 0 0 0 0 0999 V2000
  0.034584 -0.177548 0.096655 C 0 0 0 0 0 0
-0.085081 0.445186 0.982823 H 0 0 0 0 0 0
  1.037221 -0.044624 -0.309878 H 0 0 0 0 0 0
-0.703211 0.111147 -0.652091 H 0 0 0 0 0 0
-0.112391 -1.223326 0.366491 H 0 0 0 0 0 0
1 2 1 0 0 0 0
1 3 1 0 0 0 0
1 4 1 0 0 0 0
1 5 1 0 0 0 0
M END
```

**PDB** Tento formát byl vytvořen k ukládání makromolekulárních látek, ale je velmi rozšířený, jeho soubory mívají příponu `.pdb`. Má řadu možností nepotřebných pro prostou strukturu, které zde nebudou rozebírány. Na začátku je hlavička, která může obsahovat množství direktiv pro popisy jako je jméno sloučeniny či autora. Poté následují sekce pro popis multimerů a definice peptidových řetězců. Dále následuje výpis atomů, kdy atomy náležící peptidu se zapisují příkazem `ATOM` a ostatní atomy příkazem `HETATM`. V řádku za příkazem `HETATM` následuje pořadové číslo atomu, název rezidua `LIG`, kód pro vložení rezidua `1`, kartézské souřadnice, obsazenost `1.00`, teplotní faktor `0.00` a značka prvku. Pod atomy jsou vypsány vazby pomocí direktivy `CONNECT`, první číslo definuje popisovaný atom a ostatní čísla atomy k němu vázané. Direktiva `MASTER` udává počty jednotlivých direktiv, čísla `5` v uvedeném příkladu jsou pro počet `ATOM/HETATM` a `CONNECT` direktiv. Celá struktura je uzavřena direktivou `END`. Mezery musí být zachovány jako v uvedeném příkladu. [34]

```
COMPND methan
AUTHOR STEPAN SRSEN
```



```

HETATM   1  C  LIG   1      0.035 -0.178  0.097
  ↪ 1.00  0.00           C
HETATM   2  H  LIG   1     -0.085  0.445  0.983
  ↪ 1.00  0.00           H
HETATM   3  H  LIG   1      1.037 -0.045 -0.310
  ↪ 1.00  0.00           H
HETATM   4  H  LIG   1     -0.703  0.111 -0.652
  ↪ 1.00  0.00           H
HETATM   5  H  LIG   1     -0.112 -1.223  0.366
  ↪ 1.00  0.00           H
CONNECT   1   4   3   5   2
CONNECT   2   1
CONNECT   3   1
CONNECT   4   1
CONNECT   5   1
MASTER           0   0   0   0   0   0   0   0   0   5
  ↪ 0   5   0
END

```

**SDF** Samotný název toho formátu je zkratkou pro anglické *structure-data file*, což je doslova strukturní datový soubor. Tyto soubory mají příponu `.sdf` (případně `.sd`) a jedná se o obalovou strukturu nad MOL formátem, která umožňuje uložení celé databáze molekul v jednom souboru a zároveň poskytuje další možnost pro ukládání informací o molekulách. SDF formát přidává za MOL konstrukci ještě datový blok pro ukládání nestrukturních informací. Každá datová informace je složena z řádku hlavičky, řádku hodnot a prázdného řádku. Hlavička začíná znakem „větší než“ a názvem vlastnosti v ostrých závorkách. Vstup jedné molekuly je zakončen řádkou se čtyřmi dolarovými znaky. Další molekuly mohou následovat stejným způsobem. Tento formát je velmi rozšířený v obecných chemických databázích. [33]

```

methan
...
M END
> <PUBCHEM_COMPOUND_CID>
297

$$$$

```

## 4.2 Kvantové výpočetní programy

Software vyvíjený v rámci této práce potřebuje ke svému fungování kvantově chemický výpočetní program. Jelikož má být tento software primárně určen

pro akademickou sféru, základním požadavkem pro kooperující kvantový program je volná distribuce či distribuce pod akademickou licencí. Zároveň by se mělo jednat o zaběhlý a používaný program, který musí fungovat na unixových systémech, což je však v této oblasti běžná vlastnost. Díky předchozí rešerši na téma kvantových výpočetních metod je nyní možné stanovit požadavky na takový program.

Vzhledem k enormnímu počtu molekul, které jsou počítány při VS, je nutné použít relativně výpočetně nenáročné kvantové metody, což vylučuje post-HF metody. Do této kategorie tedy spadají nejméně náročné semiempirické metody, dále základní HF metoda a DFT metody. Semiempirické metody bývají standardně o tři řády rychlejší než DFT metody [35]. V závislosti na použitém výměnně-korelačním potenciálu jsou některé DFT metody rychlejší a některé pomalejší než HF metoda, řádově jsou však srovnatelné, přičemž poskytují výrazně kvalitnější výsledky [22]. Použitý kvantový program tak musí mít implementovány DFT a semiempirické metody, přičemž semiempirické metody by měli sloužit pro optimalizaci geometrií a kvalitnější TDDFT metody pro výpočet spektroskopických vlastností.

Všechny výše uvedené podmínky splňují čtyři rozšířené a běžně používané kvantové výpočetní programy: [29, 36–42]

**CP2K** CP2K je softwarový balík pro kvantovou chemii a fyziku pevných látek, jehož první verze se datuje do roku 2000. Je schopen počítat atomistické simulace pevných, kapalných, molekulárních, periodických, materiálových, krystalických i biologických systémů. Obsahuje širokou škálu DFT i semiempirických metod a klasická silová pole. Program je určený na simulaci molekulových dynamik, metadynamik, Monte Carlo a Ehrenfestových dynamik, vibračních analýz, spektroskopických vlastností, minimalizace energie a optimalizaci tranzitního stavu. CP2K je volně dostupný pod GPL licencí. [42]

**GAMESS (US)** GAMESS je program pro *ab initio* molekulární kvantovou chemii, který je dostupný pod volnou licencí. Jeho historie sahá až do roku 1981, kdy vznikl rozdělením jiného kvantového programu do dvou linií. Umožňuje optimalizace geometrie, hledání tranzitního stavu či reakční koordináty. Dále umožňuje počítat vibrační frekvence Ramanových a infračervených spekter či modelování rozpouštědla. Má implementovány některé relativistické korekce a umí fragmentovat rozsáhlé výpočty. [41]

**Firefly** Firefly, dříve pojmenovaný jako PC GAMESS, je volně dostupný *ab initio* a DFT výpočetní chemický program. Tento software se roku 1993 oddělil pod názvem PC GAMESS od programu GAMESS (US) a také s ním stále sdílí velikou část funkcionality. Autoři však uvádějí vyšší efektivitu a rychlost DFT výpočtů. [40]

**ORCA** ORCA je flexibilní a efektivní nástroj pro všestranné použití v kvantové chemii se zvláštním důrazem na spektroskopické vlastnosti. Jedná se o program dostupný pod akademickou licenci, který je vyvíjený od roku 1999. Má implementovanou celou škálu semiempirických, DFT a post-HF metod a umí také zahrnout relativistické jevy a efekty prostředí. Vývojáři kladou důraz na jednoduché použití, aby byl tento program jednoduše použitelný i mimo komunitu výpočetních chemiků. [39]

CP2K je velmi komplexní nástroj, který slouží hlavně pro simulace dynamik, což není případ zde vyvíjeného softwaru, a jeho použití je oproti ostatním programům velmi složité. GAMESS (US) a Firefly jsou funkčností i složitostí velmi podobné, přičemž Firefly má podle autorů rychlejší implementaci DFT metod. I přes svou relativně krátkou historii je ORCA z této čtveřice aktuálně nejprogressivnější a nejaktivněji vyvíjený program, který se velmi rychle dostává do popředí.

### 4.3 Zdroje dat

Tato práce se zabývá VS se zaměřením na výzkum materiálů, tomu odpovídají i relevantní datové zdroje. V aplikacích VS popsanych v kapitole 2 se jedná o nepříliš rozsáhlé organické molekuly, existuje celá řada chemických databází, kde je možné získat jejich struktury. Tyto databáze však bývají často zaměřené na určitou oblast, mívají pouze 2D struktury nebo z nich nelze data stahovat hromadně. Největším relevantním dostupným zdrojem, který zároveň poskytuje předem optimalizované 3D struktury, je databáze PubChem. [43]

PubChem vznikl v roce 2004 a je spravován americkým Národním centrem pro biotechnologické informace, které je, jako součást Národní knihovny medicíny Spojených států amerických, provozováno Národním institutem pro zdraví federální vlády Spojených států. PubChem poskytuje informace o biologické aktivitě malých organických molekul a umožňuje vyhledávání na základě různých vstupních požadavků včetně chemické podobnosti. Celý databázový systém je složen ze tří dílčích databází: PubChem Bioassay, PubChem Substance a PubChem Compound. Pubchem Substance obsahuje chemická data od jednotlivých přispěvatelů, zatímco PubChem Compound sdružuje tato data do konsolidované podoby, kdy je pro každou sloučeninu přítomen jeden ucelený výstup. PubChem BioAssay pak obsahuje data z testů biologické aktivity. [44]

Pro tuto práci je vhodná databáze PubChem Compound. Struktury molekul jsou zde uloženy ve čtyřech různých formátech: SDF, JSON, XML a ASN.1. SDF formát byl popsán v sekci 4.1, zbývající tři formáty jsou interním produktem PubChemu. Jelikož do PubChemu přispívá téměř 500 institucí, čítá tato databáze více než 93 milionů sloučenin. [45] Tyto sloučeniny jsou veřejně dostupné ke hromadnému stažení na ftp (protokol pro přenos souborů) serveru, odkud je lze stáhnout ve všech uvedených formátech vyjma formátu

JSON. [46] 3D struktury jsou vytvořeny nástrojem PubChem3D, který je počítá pomocí empirického silového pole [44]. Nejedná se tedy o výpočty vycházející ze Schrödingerovy rovnice, jako u metod v kapitole 3, ale o jednoduché parametrizované rovnice, jejichž parametry jsou získány porovnáním s experimenty či *ab initio* výsledky. V současné době má vypočtenou 3D strukturu přibližně 92 % sloučenin obsažených v PubChem Compound [45].

#### 4.4 Existující implementace virtuálního screeningu

Jak již bylo řečeno, VS je v naprosté většině případů použit pro testování účinných látek léčiv, to se odráží i v aktuálním stavu používaného software. Existuje řada komerčních i volně dostupných programů pro tento účel [47], avšak metodika v této oblasti je výrazně odlišná od přístupů používaných v materiálovém inženýrství, kam spadá náplň této práce. V posledních 4 letech však vznikají i studie se zaměřením na fotochemické vlastnosti materiálů, relativně velké procento z nich se zabývá hledáním vhodných organických materiálů pro fotovoltaické články další generace a OLED displeje (viz kapitola 2). Přestože existuje množství prací integrujících VS v rámci materiálového inženýrství a fotochemie, v této oblasti stále neexistuje žádný nekomerční veřejně dostupný software a i v komerční sféře je k dispozici zatím pouze jeden produkt. Jedná se o Materials Science Suite od firmy Schrödinger, LLC [48].

V akademické oblasti jsou používány různé metody získávání úvodní databáze molekul i různé přístupy k výpočtům jejich vlastností. Způsob získání úvodní sady molekul je vysoce závislý na řešené úloze, jelikož velikost chemického prostoru je odhadována na více než  $10^{60}$  molekul a je tak nutné ho podstatným způsobem zúžit zanesením dodatečných informací. Pokročilejší algoritmy generují molekuly na základě podobnosti strukturních motivů s již otestovanými sloučeninami nebo na základě strukturních definic chemického prostoru. Dalším přístupem jsou evoluční algoritmy, konkrétně bývají používány různé verze genetického algoritmu. Genetický algoritmus je stochastická iterativní metoda globální minimalizace, kdy se v každém kroku aplikují operátory selekce, křížení a mutace na aktuální populaci jako v evoluční biologii, v případě VS je minimalizována odchylka od požadovaných vlastností. Pro vlastní výpočet vlastností bývají používány rozličné kvantově chemické výpočetní programy, a to jak komerční, tak volně dostupné. Ve výzkumu materiálů převažuje použití semiempirických a DFT metod, v použití konkrétních metodách však neexistují obecné trendy. [1, 49–51]

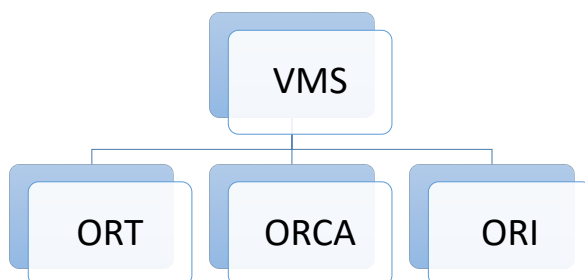
---

## Návrh

Zde vyvíjený software nese podle svého účelu název Virtual Molecular Screener (VMS). Na základě řešerše v kapitole 3 byly pro implementaci v rámci VMS zvoleny semiempirické a DFT metody, což zároveň odpovídá aktuálně používaným praktikám v materiálovém inženýrství. Následně byla jako kooperující výpočetní kvantový program zvolena ORCA, jelikož oproti ostatním programům v sekci 4.2 klade důraz na spektroskopii, což je pro VMS kýžená oblast. Zároveň je v současnosti velmi aktivně vyvíjena a její komunita se rychle rozrůstá. Bonusem je pak její jednoduché použití. Zdrojem testovacích dat byla zvolena databáze PubChem Compound (viz sekce 4.3), která má k dispozici již předem optimalizované struktury a pro účely testování VMS je možné jejich přímé použití bez další optimalizace. Z formátů dostupných v této databázi byl vybrán na základě přehledu v sekci 4.1 formát SDF.

Vzhledem k nekompatibilitě vstupního formátu kvantových programů a poskytovaných formátů v chemických databázích je nutné řešit jejich konverzi. Jelikož je však v tomto ohledu dostatečné množství volně dostupných programů, bylo by krajně neefektivní a redundantní tuto konverzi znovu implementovat. Řada nástrojů je k dispozici i online [52, 53], jejich použití však není pro databáze vhodné. Mezi volně dostupnými programy má největší oblibu Open Babel, který je dostupný pod GNU GPL licenci a podporuje více než sto chemických formátů [28, 54]. Struktury ve formátu SDF z databáze PubChem Compound tak byly převedeny pomocí Open Babel programu do XYZ formátu pro účely testování VMS.

Pokud vynecháme fázi získávání databáze molekul, kterou se VMS nezabývá, lze vlastní VS rozdělit do tří logických bloků. Nejprve je nutné nějakým způsobem nadefinovat a předat kvantovému výpočetnímu programu žádané vlastnosti a struktury molekul, poté probíhá samotný výpočet materiálových vlastností a poslední fází je vyhodnocení těchto výpočtů a vrácení příslušného výstupu. Na základě tohoto logického rozdělení je možné i reálné rozložení VMS na menší funkční celky odpovídající jednotlivým uvedeným fázím a jeden hlavní řídicí program. Rozdělení programu má řadu výhod, mezi které



Obrázek 5.1: Schéma rozdělení VMS programu na dílčí podprogramy.

patří hlavně možnost samostatného použití podprogramů a využití různých programovacích technologií.

Řídící program se nazývá VMS, program pro překlad uživatelských vstupů je nazvaný ORCA Translator (ORT), pro výpočet chemických vlastností je využíván externí kvantový program ORCA a program pro zpracování výsledků je pojmenovaný ORCA Interpreter (ORI). Schéma rozdělení je zachyceno na obrázku 5.1.

VMS program nedisponuje grafickým uživatelským rozhraním (GUI), nýbrž rozhraním příkazové řádky (CLI). Primárně je totiž určen pro unixové uživatele zvyklé na prostředí terminálu. GUI by navíc fungovalo pouze jako jednoduchý formulář, který by nepřinesl žádné zrychlení práce. Program je tedy spouštěn pomocí CLI, přičemž na vstupu obdrží soubor se strukturami molekul, vstupní soubor požadavků uživatele a výstupní adresář. Pro potřebu zadávání uživatelských požadavků byl vytvořen jednoduchý jazyk, jehož syntaxe je vysvětlena v následující sekci.

Program by měl umožňovat tři obecné typy výpočtů. Zaprvé musí umět optimalizovat geometrie do energetického minima. Zadruhé musí umožňovat výpočty absorpce a emise záření, což odpovídá výpočtům excitovaných stavů. Třetím typem je tzv. *single point* výpočet, který nemá sice přímé uplatnění pro procesy popsané v kapitole 2, ale jedná se o nejzákladnější typ kvantového výpočtu a jeho implementace vyžaduje oproti předchozím dvěma typům minimální práci navíc. Tento typ může být použit pro zjištění hodnot energie, dipólového momentu atd. a v budoucnu je možné přidání filtrů založených na *single point* výpočtech.

Výstupem programu je vždy log s názvem `orilog` a v případě optimalizace či výpočtu excitovaných stavů s filtry i databáze s názvem `output` v XYZ formátu. Při optimalizaci geometrií jsou vráceny všechny zoptimalizované geometrie. Při selekci molekul pomocí filtrů jsou vráceny molekuly splňující podmínky, případně je ještě zredukován jejich počet na hodnotu `nbest`. VMS je navržen tak, aby výstupní databázi z jednoho běhu programu bylo možné použít jako vstupní databázi dalšího běhu.

## 5.1 Uživatelský vstup programu

Pro potřeby vstupního jazyka si nejprve definujeme pojmy direktiva a filtr. Direktiva zde má význam příkazu či nastavení, které lze efektivně použít jen jednou v souboru, každý další výskyt direktivy pouze přepíše předchozí hodnotu. Naproti tomu filtry lze použít mnohonásobně a uplatňují se všechny najednou. Příkazy vstupního jazyka lze rozdělit do pěti okruhů: obecné povinné direktivy, obecné volitelné direktivy, specifické povinné direktivy, specifické volitelné direktivy a specifické volitelné filtry.

Obecné povinné direktivy musí být uvedeny v každém vstupním souboru, jinak bude vyhodnocen jako neplatný. Obecné volitelné direktivy mají význam v každém vstupním souboru, jejich uvedení však není vyžadováno, jelikož mají nastavené obvyklé přednastavené hodnoty. Specifické povinné hodnoty jsou sice povinné, ale pouze pro dané typy výpočtů, jinde nemají význam. Specifické volitelné hodnoty mají také význam pouze pro dané typy výpočtů, ovšem jejich použití není vyžadováno. Specifické volitelné filtry lze oproti specifickým volitelným direktivám efektivně použít mnohonásobně, jinak je jejich chování stejné.

I pro povinné direktivy by bylo samozřejmě možné zvolit výchozí hodnoty, ovšem nejednalo by se o žádané chování programu. Doba běhu VMS může být typicky v řádu týdnů a vynucení alespoň těch nejzákladnějších direktiv může zabránit plýtvání výpočetními prostředky na nechtěné výpočty.

### 5.1.1 Obecné direktivy a *single point* výpočty

Všechny řádky začínající znakem # jsou považovány za komentář a přeskakovány. Povinné obecné direktivy rozpoznává program dvě, je to direktiva `job` pro výběr typu výpočtu a direktiva `method` pro zvolení výpočetní kvantové metody. Typy výpočtů jsou tři a lze je vybrat pomocí klíčových slov: `single` pro *single point* výpočty, `optimize` pro optimalizaci molekul a `excited` pro výpočet excitovaných stavů. Názvy metod jsou stejné jako v programu ORCA, přičemž VMS plně podporuje HF, DFT a semiempirické metody. Kompletní seznam metod je k dispozici v ORCA manuálu [39].

Volitelné obecné direktivy jsou čtyři. Direktiva `basis` určuje použitý bázeový set, kompletní seznam je opět k dispozici v ORCA manuálu [39]. Pokud je direktiva `basis` vynechána, je vybrána vhodná báze na základě použité metody. U DFT metod bývá typicky žádaná kontrola nad výběrem báze a tím i přesností výsledku, zatímco u semiempirických metod je naopak většinou výhodné nechat výběr na kvantovém programu pro maximální efektivitu úspory výpočetního času. Direktiva `memory` nastavuje dostupnou paměť každého ORCA výpočtu v jednotkách MB, přičemž výchozí hodnota je 1024 MB. Tuto hodnotu je vhodné nastavit podle náročnosti výpočtů a použitého hardware. U výpočetních klastrů je běžně dostupná operační paměť i 8 GB na jedno výpočetní jádro. Kvantové metody vhodné pro VS však nejsou příliš

paměťově náročné a vyšší hodnoty paměti má smysl nastavovat pouze pro rozsáhlejší systémy.

Další dvě direktivy `multi` a `charge` definují spolu s geometriemi vstupní molekuly. Direktiva `multi` přijímá dvě hodnoty, přičemž první číslo musí být liché a udává multiplicitu molekul se sudým počtem elektronů a druhé číslo platí naopak pro molekuly s lichým počtem elektronů a musí být sudé. Slovo `charge` nastavuje náboj molekul a tedy se jedná o celé číslo. Výchozí hodnoty multiplicity jsou 1 respektive 2 a pro náboj je výchozí 0.

Pro *single point* výpočty nejsou vytvořeny žádné další specifické direktivy ani filtry a s použitím všech dostupných obecných direktiv má vstup pro *single point* výpočet následující tvar:

```
# General mandatory directives
job      single
method  BP86

# General optional directives
basis    def2-SV(P)
memory   2048
multi    3 2
charge   -1
```

### 5.1.2 Optimalizace geometrií

Pro optimalizaci geometrií jsou dostupné další dvě specifické volitelné direktivy. Jedná se o direktivu `iroot`, která nastavuje, pro jaký elektronový stav má být molekula zoptimalizována. Výchozí hodnota je 0, což odpovídá základnímu stavu, excitované stavy pak mají kladné celé hodnoty. Druhou direktivou je `maxdim`, ta kontroluje počet rozvojových vektorů v průběhu kvantového výpočtu. Její podstatou je zrychlení výpočtu na úkor spotřeby diskového prostoru a udává se v násobcích počtu zahrnutých excitovaných stavů. U optimalizace geometrií se jedná o násobek `iroot` a má tedy smysl pouze, pokud je `iroot` nastavený na nenulovou hodnotu. Výpočty lépe konvergují, pokud se nastaví `maxdim` na hodnotu 5-10. Základní hodnota je 10, jelikož u VS se nejedná o příliš diskově náročné výpočty, a má ji smysl měnit pouze ve specifických případech kombinace použitého hardware a velikosti počítaných molekul. Příklad vstupního souboru pro optimalizaci geometrií při použití všech dostupných direktiv:

```
# General mandatory directives
job      optimize
method  BP86

# General optional directives
basis    def2-SV(P)
```



```
memory 2048
multi 3 2
charge -1

# Job specific optional directives
iroot 1
maxdim 10
```

### 5.1.3 Výpočty excitovaných stavů

Výpočet excitovaných stavů je hlavním předmětem této práce, jelikož odpovídá absorpci/emisi záření. Tento typ má jednu specifickou povinnou direktivu, jedná se o počet excitovaných stavů `nroots`, které jsou zahrnuté do výpočtu. Výpočet excitovaných stavů má dále k dispozici čtyři specifické volitelné direktivy. Pokud je nastavena direktiva `iroot`, tak je zpracován pouze jeden excitovaný stav daný její hodnotou. Dále je možné použít direktivu `maxdim`, která má stejný význam jako u optimalizačního výpočtu, akorát zde se jedná o násobek `nroots`. Direktiva `triplets` nepřijímá žádnou hodnotu, její přítomnost však zapíná výpočet spinově zakázaných tripletových stavů (viz 2.1). Tato direktiva má smysl, pouze pokud je první hodnota `multi` nastavena na 1, což je však výchozí hodnota. Poslední volitelnou direktivou je `nbest`, která slouží k omezení počtu vrácených molekul. Má smysl pouze, pokud jsou přítomny nějaké filtry, potom vrací daný počet molekul, které nejlépe splňují zadané filtrační podmínky.

K dispozici je nyní jeden filtr s názvem `absorb`, který lze využít k selekci absorbujících či emitujících molekul. Má čtyři povinné a jeden volitelný argument a jak již bylo řečeno, je možné ho použít vícenásobně. První dva argumenty definují interval vlnových délek absorbovaného či emitovaného záření, které uživatele zajímají. Třetí a čtvrtý argument definují interval selektovaných intenzit absorpce/emise, přičemž čtvrtý argument může nést speciální hodnotu -1. Ta znamená, že horní mez intervalu je dopočítána v průběhu výpočtu podle nejintenzivněji absorbující/emituující molekuly. Pátý argument je nepovinný a udává cílenou hodnotu intenzity. Tento argument má smysl pouze při zadaném `nbest` a slouží k určení pořadí nejlepších molekul. Jeho hodnota může být přímo cílená intenzita nebo -1 pro maximální intenzitu či -2 pro střed zadaného intervalu intenzit, což je zároveň výchozí hodnota. Ukázkový vstupní soubor se všemi direktivami a různými použitími filtru `absorb` vypadá následovně:

```
# General mandatory directives
job excited
method BP86

# General optional directives
```

## 5. NÁVRH

---

```
basis    def2-SV(P)
memory   2048
multi    3 2
charge   -1

# Job specific mandatory directives
nroots   10

# Job specific optional directives
iroot    3
maxdim   150
triplets
nbest    3

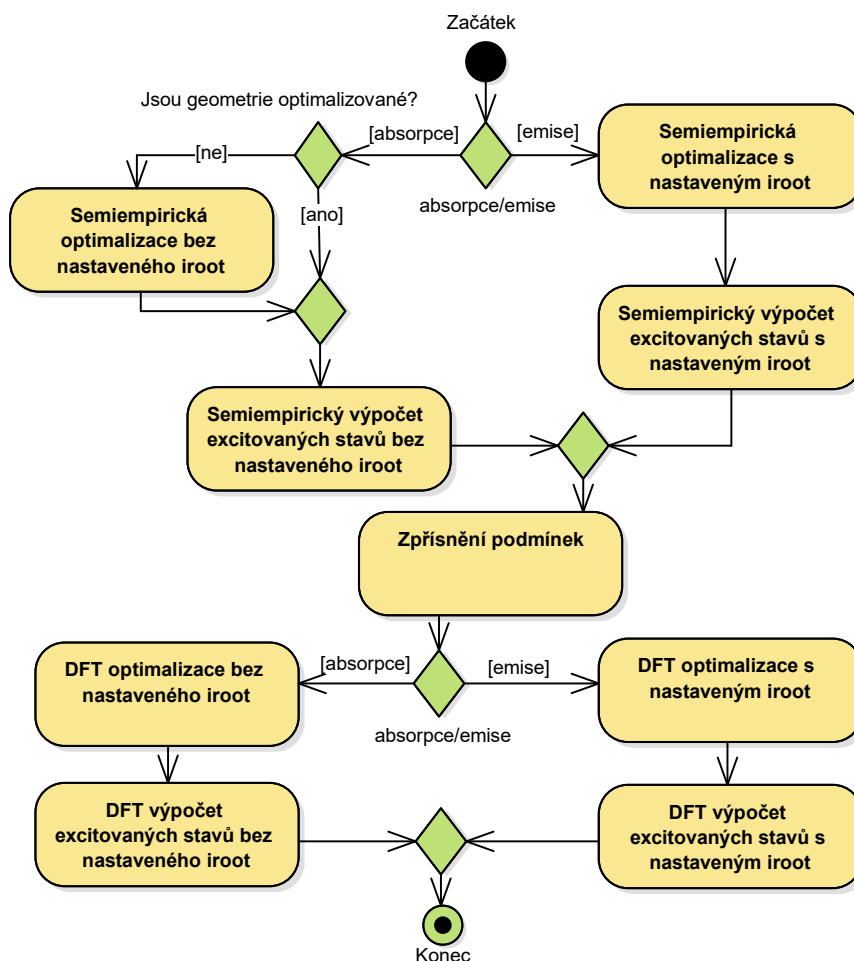
# Job specific optional filters
absorb   0 800 0.001 1.00
absorb 100 700 0.001 1.00 -1
absorb 200 600 0.002 -1.00 -2
absorb 300 500 0.003 0.08 0.25
```

## 5.2 Použití programu

VMS lze použít pro *single point* výpočty či prostou optimalizaci geometrií, jeho hlavním účelem je však selekce molekul s danými fotochemickými vlastnostmi. S výhodou lze využít iterační postup, postupně zpřísňujeme podmínky a zpřesňujeme metody. Schematické znázornění použití programu pro selekci absorbujících i emitujících molekul je zachyceno na obrázku 5.2.

### 5.2.1 Selekcce absorbujících molekul

Při selekci absorbujících molekul můžeme mít k dispozici již zoptimalizované geometrie nebo je nejdříve zoptimalizujeme pomocí semiempirické metody. Výstupní databázi použijeme jako vstup pro další běh programu, kde vyselektujeme absorbující molekuly pomocí mírně nastavených filtrů **absorb**. Tím získáme novou databázi, kterou zoptimalizujeme tentokrát s použitím DFT metody. V posledním běhu programu selektujeme molekuly DFT metodou a filtrem **absorb** s přísnějšími podmínkami, popřípadě ještě s použitím direktivy **nbest** omezíme jejich celkový počet. Tímto způsobem docílíme toho, že si špatnou semiempirickou metodou díky mírným filtrům nevyřadíme i zbytečně mnoho vhodných molekul, ale zároveň ušetříme výpočetní čas, jelikož podstatně pomalejší DFT metody používáme už na dost zredukovanou databázi.



Obrázek 5.2: Schematické znázornění využití VMS pro selekci molekul se žádanými fotochemickými vlastnostmi.

### 5.2.2 Selektce emitujících molekul

Při hledání molekul, které emitují záření v žádané oblasti a s žádanou intenzitou je proces trochu složitější. VMS umí selektovat molekuly na základě jejich fluorescence, avšak ne na základě fosforescence. Jak bylo popsáno v sekci 2.1, fosforescence je zakázaný přechod, který umožňuje pouze spin-orbitální interakce a v metodách vhodných pro VS je tak intenzita automaticky nulová. Lze si však nechat vypsát příslušné vlnové délky pomocí direktivy `triplets`.

Při hledání fluorescenčních látek používáme iterační postup jako v předchozí podsekcí a samotná selekce probíhá opět filtrem `absorb`. Geometrie však optimalizujeme do excitovaného stavu za použití direktivy `iroot`, přičemž ve většině případů je zajímavý pouze první excitovaný stav (viz kapitola 2). Při výpočtu excitovaných stavů pak používáme opět tu samou direktivu se stejnou

## 5. NÁVRH

---

hodnotou, čímž vybereme pouze přechod ze zoptimalizovaného excitovaného stavu do stavu základního. Také v tomto případě nemá většinou smysl nastavit hodnotu `nroots` větší než `iroot`. Rozdíl mezi absorpcí a fluorescencí v závislosti na zoptimalizované geometrii zachycuje také obrázek 2.4.

---

## Implementace

Jelikož je VMS určen pouze pro unixové systémy, byl pro řídicí program zvolen jazyk shellu. To umožňuje efektivní práci se soubory a filtry, která by v případě vyšších programovacích jazyků byla výrazně obtížnější. Pro podprogram ORI jako výkonnou část VMS byl vybrán jazyk C. Jeho možnosti jsou pro tento typ práce dostačující a má velmi kvalitní zoptimalizované kompilátory. Jelikož jak program ORI, tak program ORT potřebují ke své práci nejprve zpracovat uživatelský požadavek, byl pro ORT zvolen také jazyk C, aby mohly oba programy sdílet stejné metody pro zpracování vstupu. Pro samotné parsování uživatelského souboru je použit generátor syntaktických analyzátorů YACC, přičemž lexikální analyzátor je vzhledem k relativně jednoduché syntaxi implementován ručně. Parsování tak probíhá na základě nadefinované formální gramatiky. Celý program je pak sestaven pomocí utility `make`.

### 6.1 ORT

ORT program slouží k přeložení uživatelského vstupu, popsaného v sekci 5.1, na vstup pro program ORCA. Program má jeden povinný a jeden volitelný argument spuštění:

```
ort request [molecule]
```

První povinný argument definuje soubor s uživatelským vstupem. Druhý volitelný argument definuje soubor s XYZ strukturou molekuly. Podle toho, zda je druhý argument uveden, má ORT dva módy provozu. V prvním módu, kdy není uveden volitelný argument se souborem molekuly, vrací na standardní výstup obecný začátek ORCA vstupu. Pokud je soubor s molekulou uveden, pracuje ORT ve specifickém módu, kdy dopočítává ze struktury molekuly a uživatelského vstupu konkrétní multiplicitu molekuly a spolu s nábojem ji vrací na standardní výstup.

ORT zahrnuje hlavičkové soubory `atoms.h` a `request.h`. V `atoms.h` je deklarována pouze jedna funkce:

**getatom** Vstupem této funkce je zkratka atomu, výstupem pak jeho atomové číslo. To odpovídá počtu elektronů při nulovém náboji. Například `getatom("He")` vrací hodnotu 2.

Hlavičkový soubor `request.h` pak deklaruje struktury a funkce potřebné pro zpracování, ověření a uložení uživatelského vstupu:

**struct multi** Hodnota multiplicity pro sudý i lichý počet elektronů.

**struct absorb** Struktura držící právě jeden filtr `absorb`.

**struct request** Hlavní struktura pro uložení celého uživatelského vstupu.

**parsereq** Funkce pro parsování uživatelského vstupu. Jejím vstupem je soubor s uživatelským vstupem, jejím výstupem pak ukazatel na strukturu `struct request`.

**reqcheck** Vstupem této funkce je ukazatel na `struct request`, přičemž tato funkce kontroluje správnost všech uvedených hodnot.

**printreq** Funkce tisknoucí na standardní výstup obecné uživatelské požadavky v ORCA formátu. Jejím vstupem je také ukazatel na `struct request`.

**freereq** Funkce pro kompletní uvolnění struktury `struct request`.

Uvedené funkce jsou definované v souboru `request.c`, respektive funkce `parsereq` je definovaná v souboru `parse.y`, což je soubor generátoru syntaktických analyzátorů YACC. V samotném `ort.c` jsou přítomny tři funkce:

**usage** Funkce pro výpis správného použití programu.

**cm** Zkratka pro *charge* a *multiplicity*. Funkce, která na základě struktury a multiplicity nastavené uživatelem počítá celkovou multiplicitu molekuly a tiskne ji spolu s nábojem na standardní výstup. Počty elektronů jednotlivých atomů jsou získány voláním funkce `getatom`.

**main** Hlavní funkce programu, ze které jsou volány ostatní funkce. Podle počtu argumentů programu rozeznává mezi obecným a specifickým módem, jak je popsáno výše. Nejprve volá funkce `parsereq` a `reqcheck` a poté v případě obecného módu funkci `printreq` nebo v případě specifického módu funkci `cm`. Nakonec je volána funkce `freereq`.

## 6.2 ORI

Program ORI je určen ke zpracování výstupů z ORCA výpočtů na základě vstupních požadavků uživatele. Jediným vstupním parametrem programu je umístění složky s ORCA výpočty:

`ori orcadir`

Program očekává v této složce soubor s názvem `request`, který obsahuje uživatelský vstup, a ORCA logy se jménem začínajícím na „molecule.“. ORI zapisuje výsledné XYZ struktury na standardní výstup a ostatní doprovodné výsledky výpočtů na standardní chybový výstup. Stejně jako ORCA zahrnuje i ORI hlavičkový soubor `request.h`.

Ve vlastním `ori.c` jsou deklarované další tři struktury:

**struct atom** Struktura pro uložení pozice a zkratky atomu.

**struct state** Struktura pro uložení jednoho excitovaného stavu molekuly.

**struct molecule** Struktura pro uložení kompletní informace o molekule.

V `ori.c` jsou dále definované následující funkce:

**usage** Funkce pro výpis správného použití programu.

**grep** Funkce suplující stejnojmenný unixový filtr, aby nebylo nutné volat externí program.

**molnum** Funkce sloužící k získání pořadového čísla molekuly z názvu souboru.

**mkmol** Funkce pro vytvoření struktury `struct molecule` z ORCA logu.

**addmol** Realokuje pole molekul a vkládá novou strukturu `struct molecule`.

**pratom** Tiskne atom ve formátu XYZ na standardní výstup.

**prstate** Tiskne excitovaný stav na standardní chybový výstup.

**prmol\_single** Tiskne molekulu na standardní výstup u *single point* výpočtů.

**prmol\_optimize** Tiskne molekulu na standardní výstup při optimalizaci geometrií.

**prmol\_excited** Tiskne molekulu na standardní výstupy pro výpočty excitovaných stavů.

**freemol** Uvolňuje paměť po poli struktur `struct molecule`.

**print\_single** Tiskne všechny molekuly pro *single point* výpočet.

**print\_optimize** Tiskne všechny molekuly pro optimalizační výpočet.

**print\_excited** Tiskne všechny molekuly pro výpočet excitovaných stavů.

**ori\_single** Vyhodnocuje *single point* výpočty.

**read\_optimize\_line** Načítá jeden řádek optimalizované geometrie.

- read\_optimize** Načítá kompletní optimalizovanou geometrii.
- ori\_optimize** Vyhodnocuje výpočty optimalizace geometrií.
- read\_excited\_line** Načítá jeden excitovaný stav.
- read\_excited** Načítá vstupní geometrii a excitované stavy jedné molekuly.
- filter\_absorb** Vybírá molekuly na základě **absorb** filtrů.
- maxint** Počítá maximální intenzity **absorb** filtrů při hodnotě -1.
- compare** Porovnává dvě molekuly na základě spočtených kritérií.
- sort\_absorb** Řadí molekuly na základě jejich míry splnění **absorb** podmínek při zadání **nbest** v uživatelském vstupu. Molekuly jsou řazeny inteligentně podle splnění jednotlivých **absorb** filtrů, přičemž pro každý filtr je rozhodovací kritérium normalizováno na interval od 0 do 1 a jednotlivým filtrům jsou přiřazeny váhy podle obtížnosti jejich splnění.
- ori\_excited** Vyhodnocuje výpočty excitovaných stavů.
- main** Hlavní funkce programu, která nejprve zpracuje vstup uživatele a podle typu výpočtu pak volá **ori\_single**, **ori\_optimize** nebo **ori\_excited**.

### 6.3 VMS

Řídící VMS program nedělá žádné náročné výpočty a není tak nutné, aby byl napsán v kompilovaném jazyku. Vzhledem k výpočetní náročnosti celého VS je však program určen primárně pro výpočetní klastry a je tak žádána pouze přenositelnost mezi unixovými systémy. Proto je řídicí program napsán skriptovacím jazykem shell, což umožňuje snadnou práci se soubory a jednoduché použití dostupných unixových filtrů. VMS má tři povinné argumenty a řadu přepínačů:

```
vms [-h] [-a] [-d] [-f num] [-j num] [-l num] [-t dir] [-w db]
    ↪ request outdir
```

Prvním povinným argumentem je vstupní databáze molekul ve formátu XYZ. Druhým argumentem je textový soubor se vstupními požadavky (viz sekce 5.1) a třetím argumentem je výstupní složka. Přepínačů je celkem sedm a jsou následující:

- h Vypíše nápovědu a ukončí program.
- a Vytvoří archiv s celým výstupem programu.
- d Smaže výsledky, pokud jsou archivovány pomocí -a přepínače.



- f num** Pořadí první počítané molekuly v databázi.
- j num** Nastavuje počet paralelních běhů ORCA programu. U výpočetních klastrů je vhodné volit stejné číslo, které je přiřazeno procesu ve výpočetní frontě.
- l num** Pořadí poslední počítané molekuly v databázi.
- t dir** Nastavuje cestu k dočasnému adresáři, kam jsou ukládány pomocné soubory a mezivýsledky. Tento přepínač má význam obzvláště u výpočetních klastrů, které mají standardně k dispozici speciální diskový prostor vyhrazený na tyto účely.
- w** Zapíná podrobné sledování aktuálního průběhu spuštěných ORCA výpočtů a vypisuje jejich časovou náročnost.

Pro výpis varovných a informačních zpráv využívá VMS čtyři jednoduché funkce: **err**, **fatal**, **usage** a **report**. Pro načítání z databáze molekul má program dvě funkce: **skipmol** pro přeskočení jedné molekuly a **nextmol** pro načtení jedné molekuly.

VMS nejprve zpracuje a ověří použité přepínače a otestuje, zda soubory a složky předané programu mají žádané vlastnosti (čitelnost, vytváření souborů atd.). VMS také zkontroluje, zda jsou k dispozici potřebné binární soubory ORT, ORCA a ORI. Pokud je vše v pořádku, VMS načítá postupně molekuly z databáze a pomocí ORT programu vytváří ORCA vstupy. ORCA výpočty jsou postupně paralelně spouštěny a jejich počet je udržován na zadané hodnotě přepínače **-j** (výchozí hodnota je 1). Ve chvíli, kdy jsou již spuštěny ORCA výpočty pro všechny žádané molekuly, čeká program na jejich doběhnutí. Poté se volá program ORI, který výstup z ORCA logů zpracuje. Pokud je nastavený přepínač **-w**, jsou ještě v cyklu sečteny celkové časy běhů ORCA výpočtů. Nakonec program archivuje a maže výsledky, pokud jsou použity odpovídající přepínače.



---

# Testování a monitoring

VMS program byl úspěšně zkompileován a spuštěn na následujících unixových systémech:

- Debian 6.0.4 (x86\_64)
- Debian 7.11 (x86\_64)
- Ubuntu 14.04 (x86\_64)
- Ubuntu 16.04 (x86\_64)
- Gentoo 2.2 (x86\_64)

Kromě toho byl VMS úspěšně zkompileován bez přítomnosti ORCA programu na následujících systémech:

- OpenBSD 6.0 (amd64)
- OpenBSD 6.0 (armv7)
- OpenBSD 6.0 (macppc)
- SunOS 5.11 (SPARC)
- Mac OS X (x86\_64)

Všechny níže uvedené testy byly provedeny na výpočetním klastru se systémem Debian 6.0.4 a architekturou x86\_64. Testy je možné reprodukovat pomocí testovacího skriptu přiloženého k práci. Všechny testy byly provedeny s přepínačem `-w` pro sledování průběhu výpočtů. Počty molekul a metody jednotlivých testů byly vybírány podle časové náročnosti. Režijní náklady

VMS v procentech byly spočteny na základě podílu celkového času stráveného v ORCA výpočtech a času stráveného v celém VMS. Výpočet probíhal pomocí následujícího vzorce:

$$r = 100 \left( 1 - \frac{t_{ORCA, SYS} + t_{ORCA, USER}}{t_{VMS, SYS} + t_{VMS, USER}} \right). \quad (7.1)$$

Pokud není v testu stanoveno jinak, platí následující výchozí hodnoty:

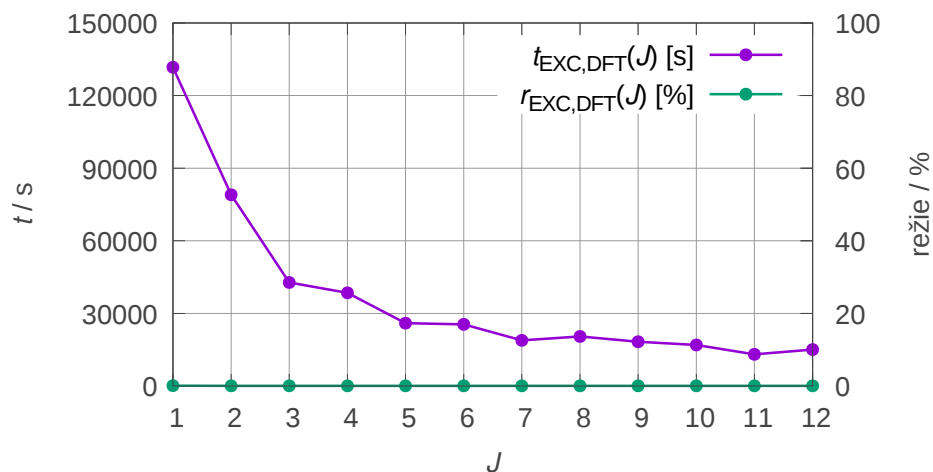
- BP86 jako výchozí DFT metoda
- ZINDO/1 jako výchozí semiempirická metoda pro optimalizaci geometrií
- ZINDO/S jako výchozí semiempirická metoda pro výpočet excitovaných stavů
- def2-SVP jako výchozí báze pro DFT výpočty
- žádná výchozí báze pro semiempirické metody
- výchozí počet excitovaných stavů 5
- ostatní parametry nenastaveny
- databáze molekul přiložená k této práci

Nejprve byla testována závislost výpočetního času a režijních nákladů VMS na počtu použitých jader. Testy byly provedeny pro výpočet excitovaných stavů s DFT a 1024 molekulami a pro výpočet optimální geometrie se semiempirickou metodou a 4096 molekulami. Výsledek je zachycen na obrázcích 7.1 a 7.2.

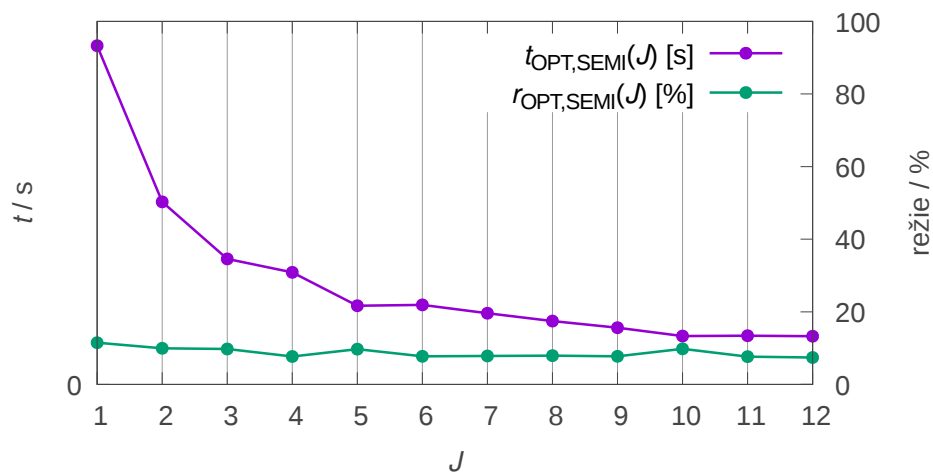
Z grafů je patrné, že v obou případech se snižuje celkový výpočetní čas nepřímou úměrou. Jedná se o správný typ závislosti, výpočet s dvojnásobným počtem jader však neodpovídá polovičnímu výpočetnímu času. K tomu přispívá zprvė režie VMS programu, ta je však u DFT výpočtu prakticky nulová. Dalším aspektem je databáze. Jelikož jsou v databázi různě veliké molekuly, mezi nimiž jsou rozdíly i několik řádů výpočetního času, dochází při relativně malém počtu molekul k jevu, kdy již většina jader má dopočteno, ale program čeká na poslední extrémně náročné molekuly. Z grafů je také patrné, že zatímco režie u velmi rychlých semiempirických metod hraje roli, u pomalejších DFT výpočtů se jedná o zanedbatelnou hodnotu.

Dále byla testována závislost celkového času a režijních nákladů na počtu molekul zahrnutých do výpočtu. Test byl proveden s 12 jádry a stejnými metodami jako výše. Výsledek je zobrazen na obrázcích 7.3 a 7.4.

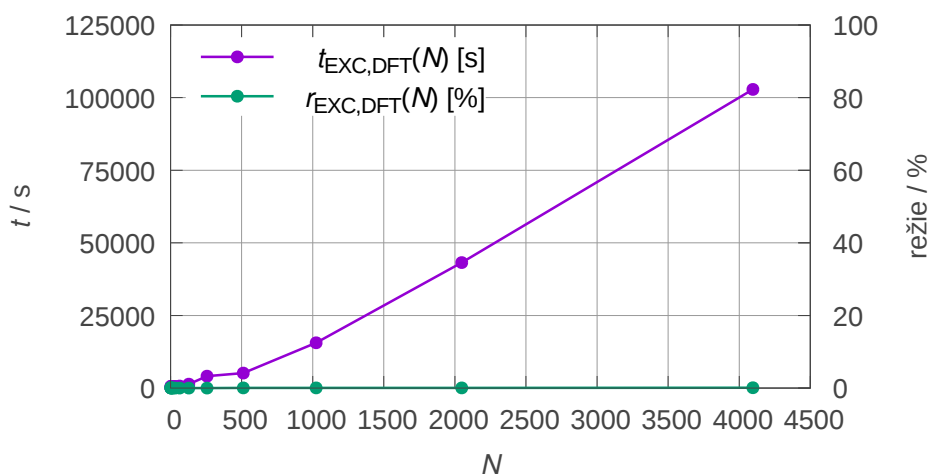
Na grafech je možné vidět, že závislost času na počtu molekul je přibližně lineární. Nelinearity v levé části grafů jsou způsobeny dvěma efekty. Zprvė dochází ke stejnému efektu s databází jako v předchozím testu, který je ovšem



Obrázek 7.1: Závislost výpočetního času a režie VMS na počtu jader pro DFT výpočet excitovaných stavů s 1024 molekulami.



Obrázek 7.2: Závislost výpočetního času a režie VMS na počtu jader pro optimalizaci geometrií semiempirickým výpočtem s 4096 molekulami.



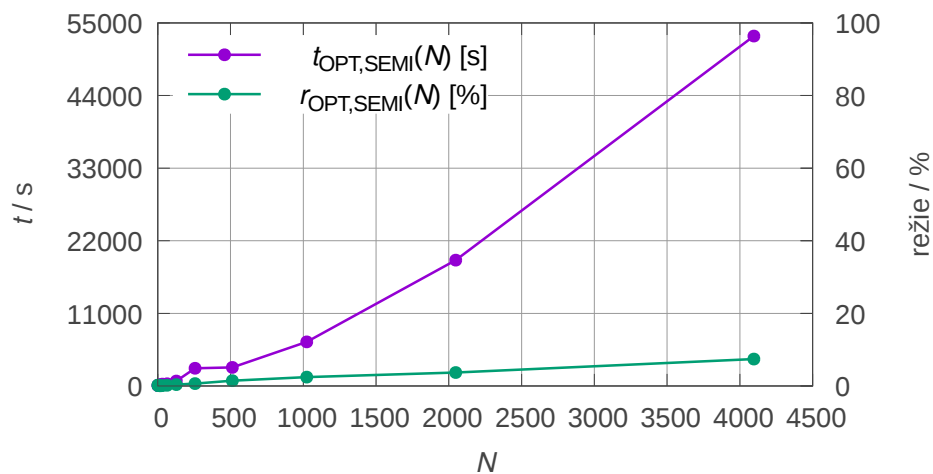
Obrázek 7.3: Závislost výpočetního času a režie VMS na počtu molekul pro DFT výpočet excitovaných stavů s 12 jádry.

ještě umocněn velmi malými počty molekul na začátku křivky. Druhý efekt souvisí s režijními náklady VMS. Ty jsou u DFT pořád prakticky nulové, ovšem u semiempirické metody lze vidět, že procento času strávené mimo ORCA výpočty lineárně roste. To je pro VS velmi negativní efekt, jelikož standardně je potřeba propočítat ještě mnohem více molekul než v těchto testech. Proto byly provedeny další testy, kdy bylo zjištěno, že stoupající tendence režie není závislá na použitém stroji, počtu jader, použité metodě a ani použité databázi. Nakonec bylo odhaleno, že tento negativní jev souvisí s nákladnou implementací přepínače `-w`, který umožňuje pokročilejší sledování spouštěných ORCA výpočtů. Pro srovnání tedy byl otestován ten samý výpočet, ovšem bez uvedeného přepínače. Přibližná oblast pro prvních 1024 molekul je na obrázku 7.5. Je zřejmé, že bez parametru `-w` je VMS zanedbatelnou zátěží oproti vlastním ORCA výpočtům i pro rychlé semiempirické metody.

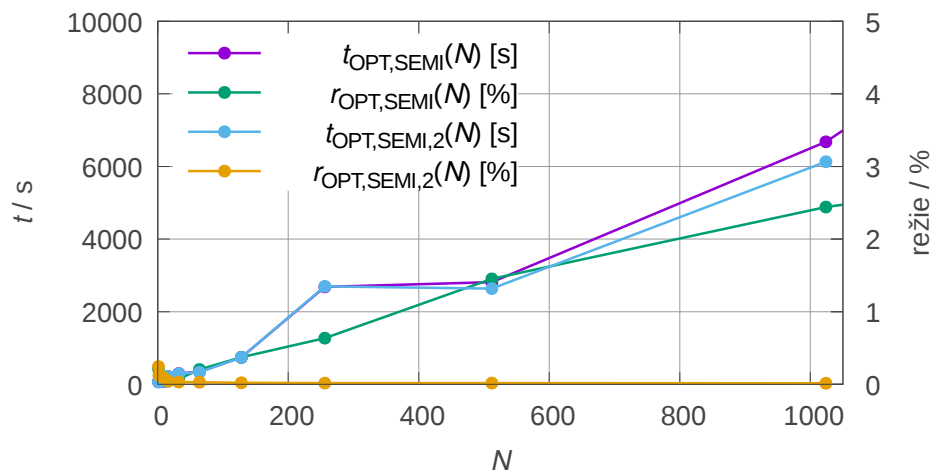
Další testy již nepopisují vlastnosti VMS programu samotného, ale pouze testují rychlosti ORCA výpočtů na základě uživatelských parametrů. Tyto testy byly provedeny s malými počty molekul a počítány s jedním jádrem.

Velmi důležitá je závislost výpočtů na počtu kontrahovaných Gaussových funkcí v bázi (viz 3.1). Báze jsou typicky k dispozici ve skupinách, kdy se postupně zvyšuje právě počet kontrahovaných Gaussových funkcí, čímž se zpřesňuje popisná schopnost báze. Závislost pro skupiny bází CC a DEF2 je pro DFT metodu s 5 molekulami zachycena na obrázku 7.6.

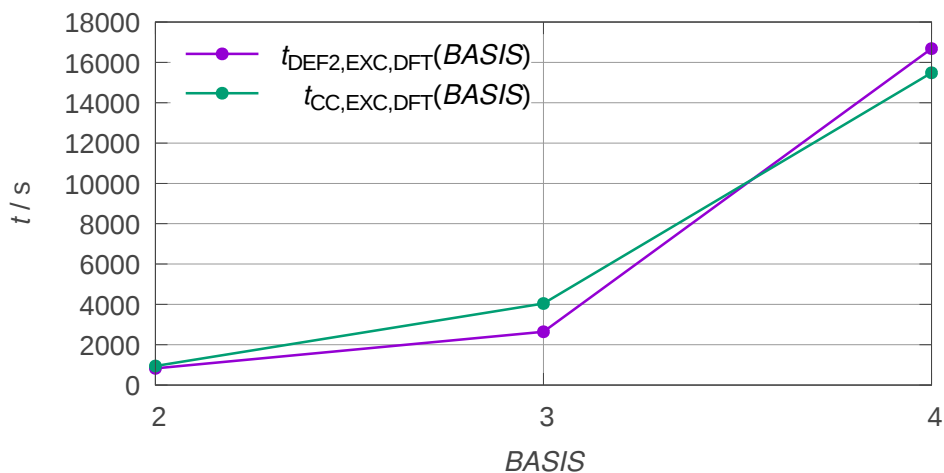
Pro více Gaussových funkcí než 4 už jsou výpočty extrémně dlouhé a výpočetně náročné. Avšak i na třech přítomných bázích v každé řadě je vidět exponenciální výpočetní složitost. Další test sleduje vliv nastavené paměti ORCA výpočtů pomocí direktivy `memory` na příkladu jedné molekuly. Z grafu na obrázku 7.7 pro DFT excitované stavy a semiempiricky optimalizované ge-



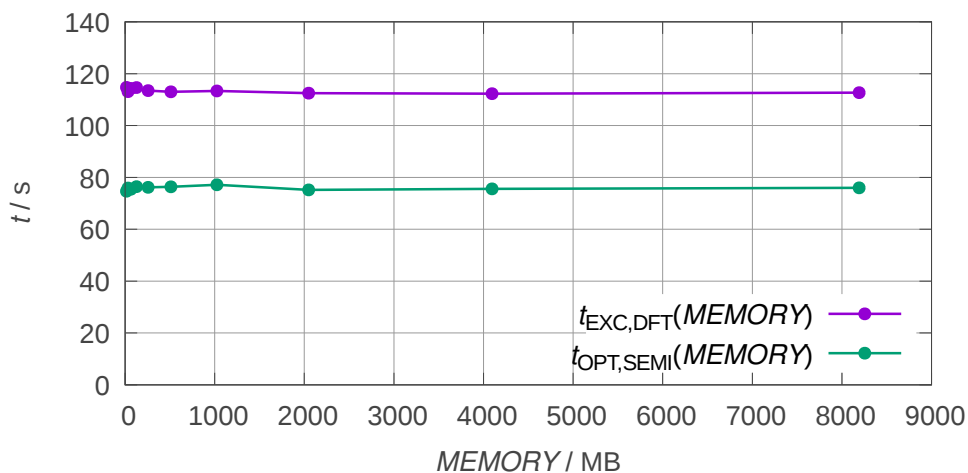
Obrázek 7.4: Závislost výpočetního času a režie VMS na počtu molekul pro optimalizaci geometrií semiempirickým výpočtem s 12 jádry.



Obrázek 7.5: Závislost výpočetního času a režie VMS na počtu molekul pro optimalizaci geometrií semiempirickým výpočtem s 12 jádry. Srovnání s výpočtem bez  $-w$  přepínače (index 2).



Obrázek 7.6: Závislost DFT výpočtů excitovaných stavů na počtu kontrahovaných Gaussových funkcí v bázi pro 20 molekul a skupiny bází CC a DEF2.

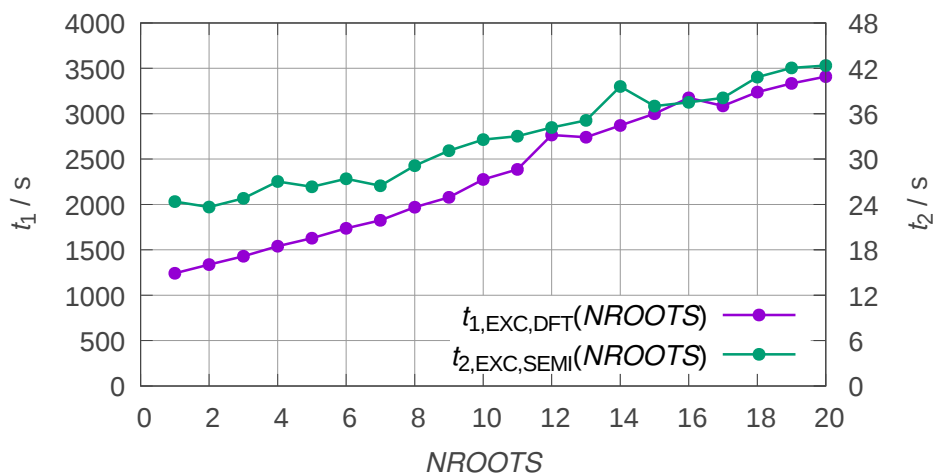


Obrázek 7.7: Závislost DFT výpočtů excitovaných stavů a semiempirických optimalizačních výpočtů na uvedené paměti. Ukázka na jedné molekule.

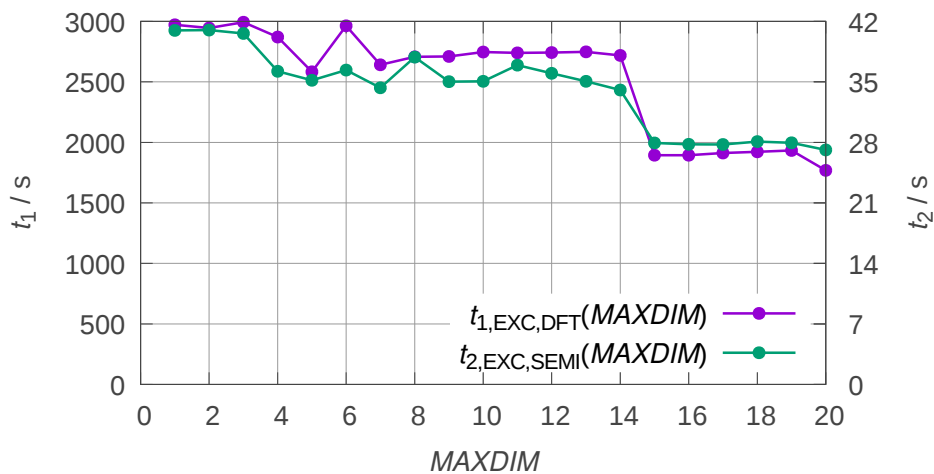
ometrie je patrné, že ani u jedné metody nemá nastavená paměť žádný vliv. Jedná se o tak malou strukturu a tak jednoduché metody, že tento parametr zde nehraje roli, jak bylo řečeno dříve v textu, direktiva `memory` je jen pro speciální případy, kdyby nastaly nějaké problémy.

Pro výpočty absorpce a emise je významné škálování s počtem excitovaných stavů a direktivou `maxdim`, výsledky těchto testů jsou zachyceny pro DFT i semiempirickou metodu na obrázku 7.8, respektive 7.9. Závislost na počtu excitovaných stavů je lineárně rostoucí, zatímco s `maxdim` sice výpočetní čas obecně klesá, ovšem jedná se o chaotický průběh.





Obrázek 7.8: Závislost výpočtů excitovaných stavů na počtu zahrnutých excitovaných stavů pro DFT a semiempirickou metodu s 20 molekulami.



Obrázek 7.9: Závislost výpočtů excitovaných stavů na parametru maxdim pro DFT a semiempirickou metodu s 20 molekulami.



---

## Závěr

Cílem této práce bylo vytvořit software pro virtuální screening fotoaktivních materiálů. Na základě rešerše byly zvoleny vhodné nástroje, metody a postupy, načež byl napsán VMS program, díky kterému je možné paralelně filtrovat molekuly, které absorbují či emitují světlo v zadaných intervalech vlnových délek a intenzit. Program byl optimalizován pro maximální možnou přenositelnost mezi unixovými systémy. VMS také obstarává práci se soubory na začátku i na konci výpočtů, nejprve si kopíruje vstupní soubory do pracovního adresáře a na konci také umožňuje archivování a mazání výsledků.

Program VMS je složen ze tří komponent, samotný VMS byl napsán v shellu, zatímco pomocné programy ORT a ORI jsou psány v jazyce C. Všechny tři komponenty byly opatřeny manuálovými stránkami. Celé VMS funguje nad chemickým kvantovým programem ORCA, který obstarává chemické jádro. Výsledný program byl testován na všechny systematické vstupní parametry, přičemž se ukázalo, že oproti samotnému kvantovému programu ORCA zvyšuje VMS výpočetní náročnost pouze zanedbatelně.

Program je psán tak, aby bylo možné doplňovat další filtry a selektovat tak molekuly na základě více kritérií, čemuž se v budoucnu budu i nadále věnovat. Na tuto práci bude navazovat bakalářskou prací kolega Jonatan Matějka, který by se měl věnovat hlavně pokročilým metodám generování databáze molekul, které jsou založeny například na základě evolučních algoritmů.



---

## Literatura

- [1] Pyzer-Knapp, E. O.; Suh, C.; Gómez-Bombarelli, R.; aj.: What Is High-Throughput Virtual Screening? A Perspective from Organic Materials Discovery. *Annual Review of Materials Research*, ročník 45, č. 1, 2015: s. 195–216, doi:10.1146/annurev-matsci-070214-020823. Dostupné z: <http://dx.doi.org/10.1146/annurev-matsci-070214-020823>
- [2] Atkins, P. W.; De Paula, J.: *Fyzikální chemie*. v Praze: Vysoká škola chemicko-technologická, vyd. 1. vydání, 2013, ISBN 8070808306;9788070808306;.
- [3] Táborský, P.: Luminiscence. 2012, [Online; cit. 2016-11-28]. Dostupné z: [https://is.muni.cz/el/1431/jaro2012/F9600/Taborsky\\_luminiscence.pdf](https://is.muni.cz/el/1431/jaro2012/F9600/Taborsky_luminiscence.pdf)
- [4] NREL: Best Research-Cell Efficiencies. 2016, [Online; cit. 2016-12-15]. Dostupné z: [http://www.nrel.gov/pv/assets/images/efficiency\\_chart.jpg](http://www.nrel.gov/pv/assets/images/efficiency_chart.jpg)
- [5] Mazzio, K. A.; Luscombe, C. K.: The future of organic photovoltaics. *Chem. Soc. Rev.*, ročník 44, 2015: s. 78–90, doi:10.1039/C4CS00227J. Dostupné z: <http://dx.doi.org/10.1039/C4CS00227J>
- [6] Brütting, W.; Frischeisen, J.; Schmidt, T. D.; aj.: Device efficiency of organic light-emitting diodes: Progress by improved light outcoupling. *physica status solidi (a)*, ročník 210, č. 1, 1 2013: s. 44–65, ISSN 1862-6319, doi:10.1002/pssa.201228320. Dostupné z: <http://dx.doi.org/10.1002/pssa.201228320>
- [7] Methane-3D-ball. 2007, [Online; cit. 2016-12-17]. Dostupné z: <https://commons.wikimedia.org/w/index.php?curid=1264890>
- [8] Jou, J.-H.; Kumar, S.; Agrawal, A.; aj.: Approaches for fabricating high efficiency organic light emitting diodes. *J. Mater. Chem. C*, roč-

- ník 3, 2015: s. 2974–3002, doi:10.1039/C4TC02495H. Dostupné z: <http://dx.doi.org/10.1039/C4TC02495H>
- [9] Klán, P.; univerzita (Brno). Katedra organické chemie, M.: *Organická fotochemie*. Brno: Masarykova univerzita, první vydání, 2001, ISBN 8021025263;9788021025264;.
- [10] Wardle, B.: *Principles and applications of photochemistry*. Chichester: Wiley, 2009, ISBN 9780470014936;0470014946;0470014938;9780470014943;.
- [11] Hilborn, R. C.: Einstein coefficients, cross sections, f values, dipole moments, and all that. *American Journal of Physics*, ročník 50, č. 11, 1982: s. 982–986.
- [12] Mertens, K.; Roth, G.: *Photovoltaics: fundamentals, technology and practice*. Chichester: Wiley, 2014, ISBN 1118634160;9781118634165;.
- [13] Espinosa, N.; Hosel, M.; Angmo, D.; aj.: Solar cells with one-day energy payback for the factories of the future. *Energy Environ. Sci.*, ročník 5, 2012: s. 5117–5132, doi:10.1039/C1EE02728J. Dostupné z: <http://dx.doi.org/10.1039/C1EE02728J>
- [14] Knapp, K.; Jester, T.: Empirical investigation of the energy payback time for photovoltaic modules. *Solar Energy*, ročník 71, č. 3, 2001: s. 165 – 172, ISSN 0038-092X, doi:[http://dx.doi.org/10.1016/S0038-092X\(01\)00033-0](http://dx.doi.org/10.1016/S0038-092X(01)00033-0). Dostupné z: <http://www.sciencedirect.com/science/article/pii/S0038092X01000330>
- [15] Kippelen, B.; Bredas, J.-L.: Organic photovoltaics. *Energy Environ. Sci.*, ročník 2, 2009: s. 251–261, doi:10.1039/B812502N. Dostupné z: <http://dx.doi.org/10.1039/B812502N>
- [16] Janssen, R. A. J.; Nelson, J.: Factors Limiting Device Efficiency in Organic Photovoltaics. *Advanced Materials*, ročník 25, č. 13, 4 2013: s. 1847–1858, ISSN 1521-4095, doi:10.1002/adma.201202873. Dostupné z: <http://dx.doi.org/10.1002/adma.201202873>
- [17] Pope, M.; Kallmann, H. P.; Magnante, P.: Electroluminescence in Organic Crystals. *The Journal of Chemical Physics*, ročník 38, č. 8, 1963: s. 2042–2043, doi:10.1063/1.1733929. Dostupné z: <http://dx.doi.org/10.1063/1.1733929>
- [18] Tang, C. W.; VanSlyke, S. A.: Organic electroluminescent diodes. *Applied Physics Letters*, ročník 51, č. 12, 1987: s. 913–915, doi:10.1063/1.98799. Dostupné z: <http://dx.doi.org/10.1063/1.98799>

- [19] Fung, M.; Li, Y.; Liao, L.: Tandem Organic Light-Emitting Diodes. *Advanced Materials*, ročník 28, č. 47, 12 2016: s. 10381–10408, ISSN 1521-4095, doi:10.1002/adma.201601737. Dostupné z: <http://dx.doi.org/10.1002/adma.201601737>
- [20] Yang, X.; Xu, X.; Zhou, G.: Recent advances of the emitters for high performance deep-blue organic light-emitting diodes. *J. Mater. Chem. C*, ročník 3, 2015: s. 913–944, doi:10.1039/C4TC02474E. Dostupné z: <http://dx.doi.org/10.1039/C4TC02474E>
- [21] Cramer, C. J.: *Essentials of computational chemistry: theories and models*. Chichester: Wiley, druhé vydání, 2004, ISBN 9780470091814;0470091827;0470091819;9780470091821;.
- [22] Lewars, E.: *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*. Springer Netherlands, 2010, ISBN 9789048138623. Dostupné z: <https://books.google.cz/books?id=5gs0v61Ec14C>
- [23] Jensen, F.: *Introduction to Computational Chemistry*. Wiley, 2007, ISBN 9780470058046. Dostupné z: <https://books.google.cz/books?id=RDIG48UcZfYC>
- [24] Hehre, W.: *A Guide to Molecular Mechanics and Quantum Chemical Calculations*. Wavefunction, 2003, ISBN 9781890661182. Dostupné z: [https://books.google.cz/books?id=Lq\\_gAAAACAAJ](https://books.google.cz/books?id=Lq_gAAAACAAJ)
- [25] Havlas, Z.: *Metody a aplikace teoretické chemie*, ročník 21. Praha: ČSAV. Ústav organické chemie a biochemie, první vydání, 1997, ISBN 8090213065;9788090213067;.
- [26] Kaplan, I.: *Intermolecular Interactions: Physical Picture, Computational Methods and Model Potentials*. Wiley Series in Theoretical Chemistry, Wiley, 2006, ISBN 9780470863336. Dostupné z: <https://books.google.cz/books?id=UX-cKCgYlW4C>
- [27] Koch, W.; Holthausen, M.: *A Chemist's Guide to Density Functional Theory*. Wiley, 2015, ISBN 9783527802814. Dostupné z: <https://books.google.cz/books?id=qm5cCwAAQBAJ>
- [28] O'Boyle, N. M.; Banck, M.; James, C. A.; aj.: Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, ročník 3, č. 1, 2011: str. 33, ISSN 1758-2946, doi:10.1186/1758-2946-3-33. Dostupné z: <http://dx.doi.org/10.1186/1758-2946-3-33>
- [29] Young, D.: *Computational chemistry: a practical guide for applying techniques to real-world problems*. New York: Wiley, první vydání, 2001, ISBN 9780471333685;0471333689;.

- [30] Rzepa, H. S.; Murray-Rust, P.; Whitaker, B. J.: The Application of Chemical Multipurpose Internet Mail Extensions (Chemical MIME) Internet Standards to Electronic Mail and World Wide Web Information Exchange. *Journal of Chemical Information and Computer Sciences*, ročník 38, č. 6, 1998: s. 976–982, doi:10.1021/ci9803233. Dostupné z: <http://dx.doi.org/10.1021/ci9803233>
- [31] Supported File Formats and Options - Open Babel v2.3.0 documentation. 2016, [Online; cit. 2016-12-15]. Dostupné z: <http://openbabel.org/docs/2.3.0/FileFormats/Overview.html>
- [32] CMLC: Chemical Markup Language - CML. [Online; cit. 2016-12-16]. Dostupné z: <http://www.xml-cml.org/>
- [33] Dynamics, N.: What is the correct format for compounds in SDF or MOL files? [Online; cit. 2016-12-16]. Dostupné z: <http://www.nonlinear.com/progenesis/sdf-studio/v0.9/faq/sdf-file-format-guidance.aspx>
- [34] wwPDB: Protein Data Bank Contents Guide. 2008, [Online; cit. 2016-12-16]. Dostupné z: [ftp://ftp.wwpdb.org/pub/pdb/doc/format\\_descriptions/Format\\_v33\\_A4.pdf](ftp://ftp.wwpdb.org/pub/pdb/doc/format_descriptions/Format_v33_A4.pdf)
- [35] Thiel, W.: Semiempirical quantum–chemical methods. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, ročník 4, č. 2, 2014: s. 145–157, ISSN 1759-0884, doi:10.1002/wcms.1161. Dostupné z: <http://dx.doi.org/10.1002/wcms.1161>
- [36] Hiessl, D.; Maurer, R.: Free chemical software information. [Online; cit. 2016-12-04]. Dostupné z: <http://www.freechemical.info/>
- [37] Linux4Chemistry. [Online; cit. 2016-12-04]. Dostupné z: <http://www.linux4chemistry.info/>
- [38] Wikipedia: List of quantum chemistry and solid-state physics software. 2016, [Online; cit. 2016-12-04]. Dostupné z: [https://en.wikipedia.org/wiki/List\\_of\\_quantum\\_chemistry\\_and\\_solid-state\\_physics\\_software](https://en.wikipedia.org/wiki/List_of_quantum_chemistry_and_solid-state_physics_software)
- [39] Max-Planck-Institute for Chemical Energy Conversion: ORCA manual. [Online; cit. 2016-12-05]. Dostupné z: <https://orcaforum.cec.mpg.de/OrcaManual.pdf>
- [40] Granovsky, A. A.: Firefly 8.2.0 manual. 2016, [Online; cit. 2016-12-05]. Dostupné z: <http://classic.chem.msu.su/gran/gamess/docs.zip>
- [41] Gordon Research Group: GAMESS Documentation. 2016, [Online; cit. 2016-12-05]. Dostupné z: <http://www.msg.ameslab.gov/GAMESS/documentation.html>



- [42] CP2K manual. 2016, [Online; cit. 2016-12-05]. Dostupné z: <https://manual.cp2k.org/>
- [43] Apodaca, R. L.: Sixty-Four Free Chemistry Databases. 2011, [Online; cit. 2016-11-08]. Dostupné z: <http://depth-first.com/articles/2011/10/12/sixty-four-free-chemistry-databases/>
- [44] Kim, S.; Thiessen, P. A.; Bolton, E. E.; aj.: PubChem Substance and Compound databases. *Nucleic Acids Research*, ročník 44, č. D1, 2016: s. D1202–D1213, doi:10.1093/nar/gkv951, <http://nar.oxfordjournals.org/content/44/D1/D1202.full.pdf+html>. Dostupné z: <http://nar.oxfordjournals.org/content/44/D1/D1202.abstract>
- [45] National Center for Biotechnology Information: The PubChem Project. [Online; cit. 2016-11-11]. Dostupné z: <http://pubchem.ncbi.nlm.nih.gov/>
- [46] National Center for Biotechnology Information: The PubChem Project FTP. [Online; cit. 2016-11-11]. Dostupné z: <ftp://ftp.ncbi.nlm.nih.gov/pubchem/>
- [47] Kellenberger, E.; Rodrigo, J.; Muller, P.; aj.: Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins: Structure, Function, and Bioinformatics*, ročník 57, č. 2, 11 2004: s. 225–242, ISSN 1097-0134, doi:10.1002/prot.20149. Dostupné z: <http://dx.doi.org/10.1002/prot.20149>
- [48] Schrödinger, LLC: Materials Science Suite. 2016, [Online; cit. 2016-12-10]. Dostupné z: <https://www.schrodinger.com/suites/materials-science-suite>
- [49] O’Boyle, N. M.; Campbell, C. M.; Hutchison, G. R.: Computational Design and Selection of Optimal Organic Photovoltaic Materials. *The Journal of Physical Chemistry C*, ročník 115, č. 32, 2011: s. 16200–16210, doi:10.1021/jp202765c. Dostupné z: <http://dx.doi.org/10.1021/jp202765c>
- [50] Kanal, I. Y.; Owens, S. G.; Bechtel, J. S.; aj.: Efficient Computational Screening of Organic Polymer Photovoltaics. *The Journal of Physical Chemistry Letters*, ročník 4, č. 10, 2013: s. 1613–1623, doi:10.1021/jz400215j. Dostupné z: <http://dx.doi.org/10.1021/jz400215j>
- [51] Pyzer-Knapp, E. O.; Simm, G. N.; Aspuru Guzik, A.: A Bayesian approach to calibrating high-throughput virtual screening results and application to organic photovoltaic materials. *Mater. Horiz.*, ročník 3, 2016: s. 226–233, doi:10.1039/C5MH00282F. Dostupné z: <http://dx.doi.org/10.1039/C5MH00282F>

## LITERATURA

---

- [52] webqc.org: Molecular formats converter. 2016, [Online; cit. 2016-12-13]. Dostupné z: <http://www.webqc.org/molecularformatsconverter.php>
- [53] University of California, Irvine: Babel: Molecule Format Converter. 2012, [Online; cit. 2016-12-13]. Dostupné z: <http://cdb.ics.uci.edu/cgi-bin/BabelWeb.py>
- [54] Open Babel. 2016, [Online; cit. 2016-12-15]. Dostupné z: <http://openbabel.org>

---

## Seznam použitých zkratek

- CLI** Rozhraní příkazové řádky
- DFT** Teorie funkcionálu hustoty
- FTP** Protokol pro přenos souborů
- GUI** Grafické uživatelské rozhraní
- HF** Hartree–Fock
- HOMO** Nejvyšší obsazený molekulový orbital
- HTS** Testování s vysokou propustností
- LCAO** Lineární kombinace atomových orbitalů
- LCD** Displej z tekutých krystalů
- LUMO** Nejnižší neobsazený molekulový orbital
- OLED** Organická elektroluminiscenční dioda
- ORI** ORCA Interpreter
- ORT** ORCA Translator
- PES** Hyperplocha potenciální energie
- SCF** Self-konzistentní pole
- TDDFT** Časově závislá teorie funkcionálu hustoty
- VMS** Virtual molecular screener
- VS** Virtuální screening
- XML** Rozšiřitelný značkovací jazyk



---

## Obsah přiloženého CD

	readme.txt.....	stručný popis obsahu CD
	bin.....	adresář s binárními soubory
	src	
	impl.....	zdrojové kódy implementace
	thesis.....	zdrojová forma práce ve formátu L <sup>A</sup> T <sub>E</sub> X
	text.....	text práce
	thesis.pdf.....	text práce ve formátu PDF
	thesis.ps.....	text práce ve formátu PS