

Střední průmyslová škola strojní   
a elektrotechnická a Vyšší odborná škola, Liberec 1, Masarykova 3

Paralelizace 3D renderování

Maturitní práce

Autor **Oliver Řezníček**

Obor **Informační technologie**

Vedoucí práce **Ing. Marek Pospíchal**

Školní rok **2022/2023**

Anotace

Práce se zabývá paralelizací procesů, vykreslením stínů a efekty následného zpracování v open-source 3D herním enginu Dotrix. Bude primárně sloužit jako výzkum a reference před implementací do samotného enginu.

Summary

This work deals with process parallelization, shadow rendering and post-processing effects in the open-source 3D game engine Dotrix. It will primarily serve as research and reference prior to implementation into the engine itself.

Čestné prohlášení

Prohlašuji, že jsem předkládanou maturitní práci vypracoval sám a uvedl jsem veškerou použitou literaturu a bibliografické citace.

V Liberci dne

Oliver Řezníček

Obsah

[Úvod 1](#_Toc125397888)

[1 Použité technologie 2](#_Toc125397889)

[1.1 Rust 2](#_Toc125397890)

[1.2 WGPU 2](#_Toc125397891)

[1.2.1 WebGPU 3](#_Toc125397892)

[1.2.2 WGSL 3](#_Toc125397893)

[1.3 Visual Studio Code 3](#_Toc125397894)

[2 Teoretická část 4](#_Toc125397895)

[2.1 Paralelizace 4](#_Toc125397896)

[2.1.1 GPU driven rendering 4](#_Toc125397897)

[2.2 Rendering 5](#_Toc125397898)

[2.3 Rendering pipeline 5](#_Toc125397899)

[2.3.1 Vertex stage 6](#_Toc125397900)

[2.3.2 Geometry stage 6](#_Toc125397901)

[2.3.3 Rasterization stage 6](#_Toc125397902)

[2.3.4 Fragment stage 6](#_Toc125397903)

[2.4 Shader 6](#_Toc125397904)

[2.4.1 Compute shader 7](#_Toc125397905)

[2.4.2 GPU commands 7](#_Toc125397906)

[2.5 Herní engine 7](#_Toc125397907)

[2.5.1 Rendering engine 8](#_Toc125397908)

[3 Praktická část 9](#_Toc125397909)

[*3.1* *Framework.rs* 9](#_Toc125397910)

[3.2 Triangle 10](#_Toc125397911)

[Závěr 11](#_Toc125397912)

[Seznam zkratek a odborných výrazů 12](#_Toc125397913)

[Seznam obrázků 13](#_Toc125397914)

[Použité zdroje 14](#_Toc125397915)

[A. Seznam přiložených souborů I](#_Toc125397916)

Úvod

Téma renderování a matematiky, která stojí za vykreslením počítačové grafiky mě vždycky zajímalo, ale nikdy jsem se mu nevěnoval tak do hloubky, jak bych chtěl. A o letních prázdninách jsem poměrně náhodou narazil na firmu Löwenware, kde jsem s odbornou pomocí vyhotovil demo *GPU driven renderování* (vykreslování řízené grafickou kartou) a toto téma mě zaujalo natolik, že jsem se rozhodl ho zpracovat jako maturitní práci.

Na začátku uvidíte jednoduchou scénu, ke které postupně budou přibývat složitější způsoby vykreslování a náročnější efekty. Tato práce tedy bude moci posloužit všem těm, kteří by chtěli více do hloubky pochopit, jak funguje renderování 2D nebo 3D grafiky. A mohu vám dopředu říct, že se znalostí WGPU budete mít velice dobré chápání toho, jak fungují dnešní moderní grafické API[[1]](#footnote-1).

# Použité technologie

## Rust

Rust je open-source, multiparadigmatický, víceúčelový programovací jazyk. Oproti jiným, jemu podobným programovacím jazykům jako je např. C nebo C++ nabízí bezpečnost paměti, která umožňuje vytvářet spolehlivé a bezpečné aplikace.

Unikátní na Rustu je hlavně to, jakým způsobem řeší bezpečný přístup do paměti. Má totiž tzv. systém vlastnictví a půjčování (ownership and borrowing), díky kterému dokáže řešit chyby, na které byste narazili v jazyce C až při běhu programu, už během kompilace. Takže se vyhnete častým chybám paměti, a to i bez použití garbage collectoru.[[2]](#footnote-2)

## WGPU

WGPU je grafické API napsané v Rustu, které umožňuje kompilaci pro více platforem díky tomu, že beží na nativních backendech jako je Vulkan, Metal, DirectX a OpenGL.

To znamená, že můžeme naši grafickou aplikaci přepnout na jiný backend a tak sestavení směřovat na úplně jinou platformu. Jinými slovy řečeno bude naše aplikace spustitelná na všech platformách. Někdo by mohl argumentovat, že aplikace napsaná ve Vulkanu nebo OpenGL bude taky multiplatformní a asi by měl pravdu, ale výhodou WGPU je že aplikace může běžet na takovém backendu, který je pro nás nebo danou platformu nejlepší.

Vývojáři by samozřejmě mohli integrovat víc grafických API do jejich herního enginu, ale to zabere spoustu času a vše se musí dopředu promyslet. Použití WGPU by jim v takovém případě práci dost usnadnilo, protože vytváří abstrakci nad všemi známými backendy.

### WebGPU

Pokud vás už podle nadpisu napadlo, že by mohla existovat určitá souvislost mezi WGPU a WebGPU, nejste na omylu. WGPU je totiž pouze Rust implementací WebGPU.

WebGPU je nadcházející moderní grafické API, které rozšíří možnosti grafiky na webu např. o compute shadery. Mluvím v budoucím čase, protože v době, kdy tuto práci píšu ještě WebGPU není plně integrované do webových prohlížečů. Zatím se objevilo pouze v testovacích verzích prohlížečů jako je Firefox Nightly nebo Chrome Canary.

### WGSL

WGSL (WebGPU Shading Language) je shader jazyk určený pro WebGPU.

## Visual Studio Code

Visual Studio Code je v současnosti asi nejznámější textový editor pro vývojáře. Vyznačuje se elegantním designem, integrovaným terminálem a mimo jiné poskytuje také nástroje pro debugging a práci s githubem. A kdyby vám v editoru přece jen něco chybělo nebo byste chtěli jiné barevné schéma, tak se můžete podívat do záložky rozšíření (extensions).

Kromě uvedených vlastností se mi líbí, jak intuitivně s ním dokáže člověk pracovat a snadno se v něm zorientovat. Pro moje účely vývoje byl tak jasnou volbou.

# Teoretická část

## Paralelizace

Paralelizace je proces, při kterém je náš program nebo výpočet rozdělen na více úloh, které běží na více procesorech. Díky rozdělení zátěže mezi více procesorů můžeme ušetřit energii a znásobit výpočetní výkon (Výkon jednoho procesoru × počet procesorů[[3]](#footnote-3)).

V případě této práce se bude jednat o přesun grafických výpočtů ze strany procesoru na grafickou kartu. Grafické výpočty jsou opakované často za sebou a dají se snadno rozložit mezi více jader (např. násobení matic). Grafická karta je tedy ideální výpočetní jednotkou pro provádění těchto operací. Má v sobě spoustu jader, které ani zdaleka nedosahují výkonu hlavního procesoru, ale jejich síla spočívá právě v počtu.

Paralelizaci chceme tedy používat hlavně v případě výpočtů, které se dají rozložit na menší, jednodušší a využijeme u nich spíš velké množství procesorů než procesor jeden s velkou frekvencí.

U paralelních výpočtů je důležitá ještě jedna věc, kterou jsem nezdůraznil, a to je souběh. Někdy potřebujeme, aby jednotlivé dílčí úlohy byly vykonávány synchronně. Předávání informací mezi úlohami může být v takovém případě dost problematické.

### GPU driven rendering

V podstatě už psaní shaderů je samo o sobě určitou paralelizací, protože tyto programy se poté spouští paralelně na grafické kartě. Paralelizace GPU driven renderování však spočívá trochu v něčem jiném.

Celý proces využívá výše uvedeného indirect draw callu, který si bere parametry z GPU bufferu místo ze samotného volání. Díky tomuto mechanismu si může grafická karta pozice a informace o následném vykreslení vzít přímo z bufferu, čímž se minimalizuje přesun dat mezi CPU a GPU. Jakým způsobem se zátěž rozloží mezi jednotlivými jádry si už obstará grafická karta.

Klíčovým bodem při designu GPU driven renderování je, že veškerá scéna (např. naše objekty) by měla být na straně GPU. Vytvoříme si velké GPU buffery (vertex + index), které se poté pošlou na grafickou kartu. Tím se snažíme vyhnout konstantnímu bindování dat mezi draw cally. Bindless design nám pomůže zlepšit výkon jak na straně procesoru, tak na straně grafické karty, kterou dokážeme s velkými draw cally lépe využít.

Moderní grafické karty milují, když jim zadáte obrovské množství práce při každém draw callu, protože tak mohou dosáhnout 100% využití. Budeme se tedy snažit, vykreslit na jeden draw call co nejvíce objektů.

## Rendering

Rendering neboli vykreslování je proces, kdy měníme naše modely objektů (jejich matematickou reprezentaci) na viditelný 2D obraz, který si můžeme zobrazit na monitoru nebo jiném zobrazovacím zařízení.

## GPU Buffers

GPU Buffer reprezentuje blok paměti, který může být použit v GPU operacích. Data jsou uložena v lineárním uspořádání, což znamená, že každý bajt může být adresován pomocí offsetu od začátku Bufferu.

### Vertex buffer

Vertex buffer je datová struktura v počítačové grafice, která uchovává vertex data jednoho nebo několika modelů v 3D prostoru. Používá se ve vertex shader stage.

Pojmem vertex data se myslí nějaký array vertexů, kde každý vertex většinou uchovává data jako:

1. Pozice – 2D nebo 3D (X, Y, Z)
2. Barva– typicky RGB
3. Normálový vektor – definuje, jak je zakřivený povrch v místě vrcholu

…

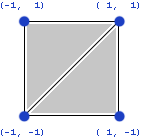
### Index buffer

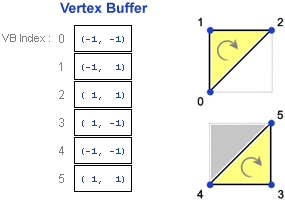
Index buffer, také známý jako index array, je datová struktura, která se v počítačové grafice používá na uchování index dat nějakého modelu.

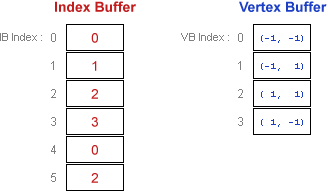
Index buffer obsahuje pole čísel, které odpovídají indexům jednotlivých vrcholu ve vertex bufferu. Tyto indexy se poté používají ke konstrukci trojúhelníků, které tvoří povrch 3D modelu.

S použitím index bufferu může GPU efektivně vykreslit stejný vertex několikrát bez duplikování dat ve vertex bufferu. Index buffer tak hraje důležitou roli ve snížení celkové paměti potřebné k uložení 3D modelu. U malých modelů není tento rozdíl v zabrané paměti tak znatelný, ale čím víc vertexů náš model má, tím výhodnější pro nás použití index bufferu bude.

#### Použití index bufferu

Pro lepší pochopení si ukážeme příklad vykreslení modelu bez a s index bufferem. Na začátku budeme mít nějaký čtverec se středem v bodě [0,0], ten vykreslíme jako dva trojúhelníky, náš primitive type tak bude *triangle list*.

V našem příkladu bude jeden vertex ve vertex bufferu reprezentován jen jako pozice v 2D prostoru. Výsledný vertex buffer by mohl vypadat takto:

Pokud do našeho řešení vykreslení čtverce zakomponujeme index buffer, tak budou naše buffery vypadat následovně:

Když budeme předpokládat, že všechna čísla zabírají v paměti 4 byty, tak v prvním případě nám bude čtverec zabírat **6×8 = 48 bytů** a v druhém případě **6×4 + 4×8 = 56 bytů**. Jak můžete sami vidět, u vykreslení dvou trojúhelníků se použití index bufferu ještě zdaleka nevyplatí, ale s jejich přibývajícím počtem se tato situace v určitém bodě zlomí a index buffer nám už bude paměť pouze šetřit.

### Uniform buffer

Uniform je něco jako globální proměnná v našem shader programu. Uniform buffer se nazývá uniform, protože se nemění v rámci jednotlivých draw callů, ale jeho hodnota je stejná (jednotná) pro všechna volání.

Do uniform bufferu si můžeme uložit data potřebná k vykreslení, která se nebudou často měnit jako např. projection nebo view matici.

### Storage buffer

Uniform buffery jsou vhodné pro malá data určená pouze ke čtení. Když ale budete chtít mít v shaderu data o neznámé velikosti nebo data, která chcete měnit. Na to se používají storage buffery. Vytvářejí se stejným způsobem jako uniform buffery. Vlastně i fungují hodně podobně, mají jen jiné vlastnosti jako větší maximální velikost, a že se do nich dá zapisovat ze shaderu.

Storage buffery jsou obvykle o něco pomalejší než uniform buffery, ale můžou být mnohem větší. Pokud např. chcete nacpat celou scénu do jednoho bufferu, budete je muset použít.

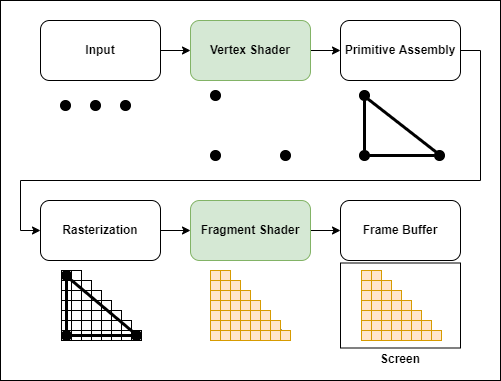
### Indirect buffer

Indirect buffer se používá v indirect draw callu, kde nahrazuje funkci parametru. Při volání indirect draw commandu si pouze zvolíme odkud z indirect bufferu chcete kreslit a grafická karta z tohoto místa spustí draw command.

Má jasně definovanou strukturu, která se skládá z indirect draw commandů. Ty v sobě obvykle obsahují informace jako odkud máme začít kreslit z index/vertex bufferu, počet vertexů k vykreslení nebo kolik instancí daného modelu chceme vykreslit.

Díky použití indirect bufferu můžeme snížit počet draw callů potřebných k vykreslení naší scény. Bude nám tak užitečný v situacích, kdy potřebujeme vykreslit velké množství objektů (příkladem můžou být particly v particle systému).

## Rendering pipeline

Rendering pipeline je sekvence kroků, kterými musí naše objekty ve scéně projít, než se vykreslí na obrazovku. Typicky se dělí na tyto uvedené na obrázku:

Obrázek - WGPU render pipeline

Zeleně vybarvené části jsou programovatelné. Kód, kterým můžeme ovlivnit chování v těchto krocích, se píše do **shaderu**.

### Vertex stage

Vertex stage je zodpovědná za zpracování dat jednotlivých vertexů. To většinou zahrnuje transformaci souřadnic z world space do view space[[4]](#footnote-4). Typicky zde provádíme maticové operace jako např. násobení.

### Geometry stage

Po vertex stage následuje geometry stage nebo také tesselation stage. Ta je zodpovědná za sestavování primitiv (např. trojúhelníků) z vertexů.

### Rasterization stage

Rasterization stage určuje pixely, které překrývají daná primitiva, abychom mohli vypočítat barvu všech překrytých pixelů.

### Fragment stage

Ve fragment stage spočítáme pro každý pixel překrytý primitivem očekávanou barvu. Barva se většinou předává rovnou s vertexy nebo ji můžeme namapovat z textury. Také zde můžeme provádět výpočet světla, abychom rozlišili, kolik světla dopadá na určité pixely.

Výstupem této stage je už konečný frame buffer.

## Shader

Shader je počítačový program, který slouží k popsání programovatelných částí rendering pipeline. K tvorbě shaderů se používají speciální programovací jazyky tzv. shader jazyky. V našem případě budeme používat jazyk **WGSL**.

Nejčastěji se budete potkávat s těmito typy shaderů:

1. **Vertex shader**
2. **Fragment shader**
3. **Compute shader**

Pro co se používá vertex a fragment shader je popsáno výše u jednotlivých částí rendering pipeline, takže tyto dva vynechám a spíš popíšu, k čemu slouží compute shader.

### Compute shader

Když budete potřebovat udělat nějaký dodatečný výpočet s použitím grafické karty, tak přichází na řadu compute shader. Pomocí něj můžeme urychlit obecné algoritmy spuštěním na grafické kartě. Jedním z příkladů použití compute shaderů může být generování meshů nebo vytváření particle systémů.

Proč tedy není v render pipeline něco jako compute stage? Compute shader je taková pomocná ruka, kterou při renderování občas budeme potřebovat, ale se samotným procesem render pipeline nemá nic společného. Je totiž součástí compute pipeline[[5]](#footnote-5).

### GPU commands

1. **Draw command** – spouští render pipeline
   1. **Draw** – kreslí primitiva
   2. **Draw indexed** – kreslí indexovaná primitiva
   3. **Draw indirect** – kreslí primitiva z dat, které oproti obyčejnému draw callu jsou předávány v GPU Bufferu
   4. **Draw indexed indirect** – kombinuje dva předchozí draw cally dohromady – je to tedy indexovaný indirect draw
2. **Dispatch command** – spouští compute pipeline

## Herní engine

Herní engine je vývojové prostředí speciálně navržené tak, aby pomáhalo vývojářům vytvářet hry efektivněji. Poskytuje vývojářům sadu nástrojů, které umožňují tvorbu počítačových her bez nutnosti stavět vše od začátku. Typicky herní engine poskytuje nástroje pro vykreslování 2D a 3D grafiky, simulaci fyziky, skriptovací jazyk a různé další funkce.

Zjednodušeně se tak vývojáři mohou více soustředit na vytváření herních mechanismů a ostatní věci nechat na enginu.

### Rendering engine

Rendering (vykreslovací) engine je zodpovědný za vykreslení jednotlivých objektů, stínů a dalších věcí v naší scéně. Všechno, co v počítačové hře vidíte na obrazovce je tedy (mimo spousty práce, kterou museli hře věnovat herní vývojáři a designéři) zásluhou vykreslovacího enginu[[6]](#footnote-6).

# Praktická část

Celý projekt je rozdělen na několik příkladů, které popisují, jak se projekt postupně vyvíjel. Příklady nám tak pomáhají lépe pochopit jednotlivé problémy a jejich konkrétní implementaci.

## *Framework.rs*

Celý návrh tohoto projektu vychází z WGPU githubu a stěžejní pro nás bude soubor *framework.rs* nacházející se ve složce examples.

*Obsah obrázku text

Popis byl vytvořen automaticky*Na začátku souboru se nachází *trait[[7]](#footnote-7)* s názvem *Example*, který představuje náš konkrétní příklad. Ten budeme pro ulehčení používat vždy při vytváření nového examplu.

Následují tři podprogramy *setup*, *start* a *run[[8]](#footnote-8)*.

1. Podprogram *setup* vytváří nové okno v operačním systému a inicializuje WGPU.
2. Druhý podprogram v pořadí s názvem *start* spouští naší aplikaci a vytváří event loop našeho okna
3. Obsah obrázku text

   Popis byl vytvořen automatickyTřetí podprogram už jen dává ty dva první dohromady. Takže spustí asynchronně *setup* a výsledek z něj předá podprogramu *start*.

Program *framework.rs* tedy tvoří vlastně gro celé naší aplikace. Při vytvoření nového příkladu si vždy nadefinujeme novou strukturu implementující trait *Example* a ten pouze předáme jako generický parametr funkci *run*. Tímto předáním se tak vždy spustí naše nově nadefinované chování v podprogramech *init*, *resize*, *update* a *render*.

## Triangle

Úplně prvním z příkladů v tomto projektu je jednoduchý trojúhelník, na kterém si vysvětlíme proces vytváření render pipeline.

Závěr

Seznam zkratek a odborných výrazů

Engine

Softwarový engine je software, který poskytuje základní funkce pro ostatní software. Je to vlastně takový „motor“, který pohání nějakou aplikaci a zpravidla je jádrem celé její funkčnosti. Příkladem může být databázový engine, vykreslovací engine nebo herní engine.

Vertex

Vertex je bod v prostoru.

Mesh

Mesh je kolekce vertexů, hran a stěn, které dohromady definují nějaký tvar nebo těleso. Příkladem takového tělesa může být koule.

Seznam obrázků

[Obrázek 1 - WGPU render pipeline 8](file:///C:\Users\Oliver%20Řezníček\Documents\Olas\Škola\Maturita\Maturitní%20práce\paralelizace_3d_renderovani\docs\Paralelizace_3D_renderovani.docx#_Toc126003839)

Použité zdroje

1. **sotrh.** Textures and bind groups. *Learn Wgpu.* [Online] https://sotrh.github.io/learn-wgpu/beginner/tutorial5-textures/.

2. **WebGPU Shading Language. *W3C.* [Online] https://www.w3.org/TR/WGSL/.**

**3. WGPU documentation. *Docs.rs.* [Online] https://docs.rs/wgpu/latest/wgpu/.**

**4. *The Rust Programming Language - book.* [Online] https://doc.rust-lang.org/book/.**

**5. Paralelní výpočty. *Wikipedie.* [Online] https://cs.wikipedia.org/wiki/Paraleln%C3%AD\_v%C3%BDpo%C4%8Dty.**

**6. AZAT. History of Rust Programming Language. *TechnoSuggest.* [Online] 28. Červenec 2022. https://technosuggest.com/history-of-rust-programming-language/.**

**7. Rust (programovací jazyk). *Wikipedie.* [Online] https://cs.wikipedia.org/wiki/Rust\_(programovac%C3%AD\_jazyk).**

**8. GPU Driven Rendering. *Vulkan Guide.* [Online] https://vkguide.dev/docs/gpudriven/gpu\_driven\_engines/.**

**9. Microsoft. Rendering from Vertex and Index Buffers (Direct3D 9). *Windows App Development.* [Online] 1. 6 2021.**

1. Seznam přiložených souborů

Na přiloženém datovém nosiči se nacházejí následující soubory a složky:

* **Aplikace** – zdrojové kódy

1. Příkladem takového moderního grafického API může být Vulkan. [↑](#footnote-ref-1)
2. Jednotlivými koncepty a syntaxem jazyka Rust se ve své práci nebudu dopodrobna zabývat, takže pokud byste se chtěli o Rustu dozvědět víc a třeba si v něm i zkusit napsat vlastní program, určitě bych pro začátek doporučil elektronickou knihu Rust book, kterou mám uvedenou ve zdrojích. [↑](#footnote-ref-2)
3. Procesor možná není v tomto případě úplně vhodný termín, protože se může jednat pouze o vlákno nebo jádro našeho procesoru v PC. Vhodnější by byl asi termín výpočetní jednotka, ale pro jednoduchost necháme procesor. [↑](#footnote-ref-3)
4. World space je náš 3D svět do kterého jsme si umístili nějaké objekty. Oproti tomu view space je prostor naší 2D obrazovky, kam chceme naše objekty přetransformovat. [↑](#footnote-ref-4)
5. Tady je důležité podotknout, že všechny tyto typy shaderů můžeme psát do jednoho **WGSL** souboru. [↑](#footnote-ref-5)
6. Důležitý je pro nás tento termín hlavně proto, že vykreslovací engine bude jedinou částí herního enginu, kterou se tato práce bude zabývat. [↑](#footnote-ref-6)
7. Trait v Rustu je něco jako abstraktní třída nebo interface. Umožňuje přidat chování do třídy bez nutnosti použití dědičnosti. V našem případě nám definuje nějaké metody a mohl by i vytvářet jejich defaultní implementaci. [↑](#footnote-ref-7)
8. Z důvodu délky jednotlivých kódů vám ukážu pouze podprogram *run*. [↑](#footnote-ref-8)