

Konzeption und Implementierung eines Unified Rendering Frameworks mit modernen GPU Computing APIs

Bachelorarbeit

zur Erlangung des Grades eines Bachelor of Science (B.Sc.)
im Studiengang Computervisualistik

vorgelegt von
Markus Schlüter

Erstgutachter: Prof. Dr.-Ing. Stefan Müller
(Institut für Computervisualistik, AG Computergraphik)

Zweitgutachter: Dipl.-Inform. Dominik Grüntjens

Koblenz, im Mai 2011

Erklärung

Ich versichere, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Ja Nein

Mit der Einstellung der Arbeit in die Bibliothek bin ich einverstanden.

Der Veröffentlichung dieser Arbeit im Internet stimme ich zu.

.....
(Ort, Datum)

.....
(Unterschrift)

Todo list

evtl. beispielschema erstellen für zwei klassische transformationsklassen und adapter vs. unified transformation	2
Zitat einfügen? Stefan Müller, PCG? ;)	2
Muss ich das irgendwie belegen? Ich wüsste nicht, wie/wo ;(.	4
referenz auf Peschel, evtl weitere eigenschaften wie curl und sicherstellung der inkompressibilität	4
referenz auf gpu-gems-gridbased-zeug, evtl jos stam	4
referenz auf thomas steil und GPU gems	5
ref auf peschel	5
'indirekt' erläutern? transform feedback buffer, scattering über gl_Position und ein-pixel point rendering, siehe DA von Sinje Thiedemann?	5
Wo zum Henker soll ich dazu nun wieder ne Referenz herauskramen? Ich habe leider keine idee, außer zig allgemeine Papers und Übcher zu überfliegen Zeitproblem!... das ist der nachteil, wenn man so viel recherchiert hat, zunächst mal fürs Verständnis und oder oder den ganz groben überblick, und nicht für eine umfassende wissenschaftliche Dokumentation: All den Input, der nur als Hintergrundwissen dienen sollte, jetzt noch wieder zu finden... :(Wie kann ich mit diesem allgemeinen Problem umgehen?	5
Referenz eines guten Artikels zu GPGPU finden;	5
ref auf späteren abschnitt oder OpenCL spec	6
hier bin ich mir nicht mal sicher, ob das stimmt; meine Erinnerung ist nur sehr vage, diese Fragestellung wurde in meiner Literatur höchstens ganz marginal behandelt; Was tun? Aussage weg lassen, mich tod- suchen, oder diese gewagte These nach bestem Wissen und Gewissen stehen lassen?	6
hier specs und refs zu PCIe 2.0 anbringen? eher nicht, oder?	8
Beleg? Wie war das mit Folding@Home?	10
referenz finden, bin mir da eher unsicher	11
explizite Danksagung?	11
evtl treffenderen Ausdruck finden: Scanline-basiert oder was auch immer	15
referenz zu PCIe-Flaschenhals-stuff	58
diesen klumbatsch in form bringen, mit bildern anreichern etc pp	59
überlegen, ob ich aus Interesse nicht noch weiter in die Richtung recherchieren sollte, da ich nach meiner Implementierung erst so richtig beeindruckt von dem Verfahren war (ich habe im Internet noch keine Fluid-Demo gefunden, die ebenfalls SPH implementiert; ok., ich hab auch nicht gesucht ;)), und gerne mehr über die Hintergründe verstehen würde... problem, wie immer: Zeitdruck ;(.	62

Inhaltsverzeichnis

1 Einleitung	1
1.1 Motivation	2
1.1.1 „Unified Rendering Engine“	2
1.1.2 Fluidsimulation	4
2 Überblick	7
2.1 Vision	7
2.2 Paradigmen	7
2.3 Begriffe	8
2.4 Schwerpunkte	10
3 Systemarchitektur	15
3.1 Dependencies	15
3.2 Klassendiagramm	18
3.3 BasicObject und Memory Tracking	18
3.4 AmendedTransform	22
3.5 Die <i>Unified Rendering Engine</i> (URE)	25
3.6 Das UserInterface-Paket	25
3.7 Das Simulator-Paket	26
3.8 Der SimulationResourceManger	28
3.9 Das Scene-Paket	28
3.10 Das WorldObject	30
3.11 Material	31
3.12 Die Buffer-Abstraktion	33
3.12.1 Die API-bezogenen internen Buffer-Operationen	38
3.12.2 PingPongBuffer	39
3.13 Geometry	40
3.13.1 BufferBasedGeometry	41
3.13.2 VertexBasedGeometry	42
3.13.3 InstancedGeometry	43
3.14 Das MPP-Paket	43
3.14.1 MPP und ParallelComputeManager	43
3.14.2 Shader und ShaderManager	45
3.14.2.1 Probleme	50
3.14.3 CLProgram und CLProgramManager	52
3.14.3.1 CLKernelArguments	53
3.15 Status der Implementierung am Ende der BA	53
4 Simulation	58
4.1 Die visuelle Simulationsdomäne	58
4.1.1 Der LightingSimulator	58
4.1.2 Die Lighting Simulation Pipeline Stages	58

4.1.3	VisualMaterial	58
4.1.4	Camera	58
4.1.5	LightSource und LightSourceManager	58
4.1.6	RenderTarget	58
4.1.7	ShaderManager	58
4.1.8	genutzte moderne OpenGL- Features	58
4.1.8.1	Hintergrund:Batching	58
4.1.8.2	Uniform Buffers	58
4.1.8.3	Instancing	59
4.1.8.4	Hardware Tesselation	59
4.1.9	Ablauf	59
4.1.10	Implementierte Effekte	59
4.2	Die mechanische Simulationsdomäne	62
4.2.1	Fluidsimulation	62
4.2.1.1	Grundlagen	62
4.2.1.1.1	Die Navier-Stokes-Gleichungen	62
4.2.1.1.2	Grid-basierte vs. Partikelbasierte Simulation .	62
4.2.1.1.2.1	Die zwei Sichtweisen: Lagrange vs. Euler	62
4.2.1.1.3	Smoothed Particle Hydrodynamics	62
4.2.1.2	Verwandte Arbeiten	62
4.2.1.3	Umfang	63
4.2.1.4	Algorithmen	63
4.2.1.4.1	Work Efficient Parallel Prefix Sum	63
4.2.1.4.2	Parallel Radix Sort und Stream Compaction .	63
4.2.1.5	Ablauf	63
4.2.1.6	Hardwarespezifische Optimierungen	63
5	Ergebnisse	64
6	Fazit	68
7	Ausblick	69
A	Listings zur Buffer-Abstraktion	71

Abbildungsverzeichnis

1	Gegenüberstellung von Verwendung und Begrifflichkeiten von klassischen Engines und einer Unified Engine	8
2	Klassendiagramm des Gesamtsystems, Teil 1	19
3	Klassendiagramm des Gesamtsystems, Teil 2	20
4	Grobes Beispiel von einer Simulation und den zugehörigen Daten-Abhängigkeiten	27
5	Der Begriff <i>Material</i> mit einer physikalisch orientierten Bedeutung: sämtliche Eigenschaften von Materie außer ihrer geometrischen Form. (Bild entnommen aus [APS00])	32
6	Sequenz-Diagramm der Shader-Generierung	47
7	Artefakte bei naiver Tessellation-Level-Berechnung mit Camera-Viewspace bei der Shadow Map Generation	50
8	Klassendiagramm der OpenCL-relevanten Klassen inkl. Skizze der Call-Hierarchie des Build-Prozesses sowie dem Kernel-Launch	55
9	Überblick über die prototypische Szene	57
10	Die Extremitäten des Raptor-modells; Oben: tesseliert, texturiert, normal mapped, environment mapped, beleuchtet von fünf Lichtquellen, shadow mapped von einer Lichtquelle; Unten: wie oben, nur ohne Tessellation	64
11	Füße des Raptor-Modells, mit verschiedenen Permutationen aktiverter Shader-Features	65
12	seffe	66
13	Fuß des Raptor-Modells, Gegenüberstellung des Detail-Grades mit und ohne Tessellation	67

Tabellenverzeichnis

1	Gegenüberstellung der Zwecke einer generischen Szenen-Repräsentation und einer Beschleunigungsstruktur	29
2	Verschiedene Buffertypen und ihre Verfügbarkeit in verschiedenen Kontexten Legende: ✓ → nativ unterstützt; ○ → kompatibel; ✘ → nicht unterstützt; ? → Unterstützung abhängig von weiteren Parametern, s. Tabelle 3;	37
3	Verschiedene Texturtypen und ihre Kompatibilität zu bestimmten Features Legende: ✓ → unterstützt; ○ → In OpenCL unterstützt, aber nicht vom Framework; ○ → nur in OpenGL 4 unterstützt, daher wegen Kompat. zu GL 3 nicht vom Framework unterstützt; ✘ → nicht unterstützt;	37

4	Status der Implementation zum Zeitpunkt der Abgabe der Ausarbeitung (englisch wegen vielen Fachbegriffen) Legende: ✓ → implementiert; ○ → zu weiten Teilen programmiert, jedoch noch nicht integriert/ausgeführt/getestet; ○ → detailliert Konzipiert, jedoch nicht programmiert; ✗ → im System langfristig konzipiert, jedoch nicht programmiert; - → Unterstützung nicht geplant	56
5	Status der Implementation zum Zeitpunkt der Abgabe der Ausarbeitung (Fortsetzung)	57

1 Einleitung

Im Rahmen dieser Bachelorarbeit wurde der Frage nachgegangen, inwiefern eine sogenannte „Unified Rendering-Engine“, welche verschiedene Simulationsdomänen vereint, einen Mehrwert darstellen kann gegenüber dem klassischen Ansatz, z.B. gesondert sowohl eine Graphik- als auch eine Physik-Engine zu verwenden, die zunächst einmal keinen Bezug zueinander haben;

Hierbei wurde besonderer Wert auf die Verwendung moderner GPU-Computing-APIs gelegt, namentlich auf OpenGL3/4 und OpenCL.
Da bei diesem ganzheitlichen Thema eine vollständige Implementierung einer solchen vereinheitlichten Engine unmöglich war, konnte nur ein Bruchteil der Konzepte implementiert werden;

Dieser Umstand war von Vornherein bekannt, und die Versuchung ist stark, wie in einer Demo die schnelle Realisierung eines Feature-Sets einer konsistenteren, aber zeitaufwändigeren und zunächst karger wirkenden Implementierung vorzuziehen. Dieser Versuchung wurde versucht, nur dort nachzugeben, wo die negativen Auswirkung auf die Konsistenz des Gesamtsystems lokal bleiben, und so nicht „Hacks“ sich irreversibel durch das gesamte System ziehen.

Letztendlich wurden exemplarisch für die Nutzung in der visuellen Simulationsdomäne einige gängige visuelle Effekte einer Graphik-Engine implementiert, wie Shadow Mapping, Normal Mapping, Environment Mapping, Displacement Mapping und dynamisches LOD. Es wurden moderne OpenGL- und Hardware-Features wie Instancing, Uniform Buffers und Hardware-Tesselation verwendet. Schwerpunkt war hier der Einsatz einer Template-Engine, damit

- Boilerplate-Code in den Shadern vermieden wird und
- Effekte beliebig (sinnvoll) nach Möglichkeit zur Laufzeit miteinander kombinierbar sind

Mehr dazu in Kapitel 4.1

In der mechanischen Simulationsdomäne wurde eine partikelbasierte Fluidsimulation mit OpenCL auf Basis von Smoothed Particle Hydrodynamics implementiert. Mehr dazu in Kapitel 4.2.

Das System trägt den Namen „Flewnit“, eine bewusst nicht auf den ersten Blick erkennbar sein sollende¹ Kombination der Worte „Fluid“, in

¹Es soll der generische Ansatz des Frameworks nicht in den Hintergrund gedrängt werden.

Anspielung auf den ursprünglichen Zweck einer Bibliothek zur Fluidsimulation und „Unit“, in Anspielung auf „Unity“-„Einheit“. Zufälligerweise ist das *Nit* auch noch die englische Einheit für die Leuchtdichte, $\frac{Cd}{m^2}$.

1.1 Motivation

Ursprünglich als Arbeit zur Implementierung einer Fluidsimulation geplant, wurde bald ein generischer, eher softwaretechnisch orientierter Ansatz verfolgt, der jedoch die Implementierung einer Fluidsimulation als mittelfristiges Ziel hatte;

1.1.1 „Unified Rendering Engine“

Der Wunsch nach einer „Unified Rendering Engine“ erwächst aus eigener Erfahrung der Kopplung von Physik- und Graphik-Engines, namentlich der Bullet Physics Library² und der OGRE Graphik-Engine³. Diese Hochzeit zweier Engines, die jeweils für verschiedene „Simulationsdomänen“ zuständig sind, bringt gewissen Overhead mit sich, da Konzepte wie Geometrie und ihrer Transformationen unterschiedliche Repräsentationen bzw. Klassen zugrunde liegen; Hierdurch wird die gemeinsame Nutzung beider Domänen von Daten wie z.B. Geometrie nahezu unmöglich; Ferner müssen für eine die beiden Engines benutzende Anwendung diese Klassen mit ähnlicher Semantik durch neue Adapterklassen gewrapped werden, um dem Programmier der eigentlichen Anwendungslogik den ständigen Umgang mit verschiedenen Repräsentationen und deren Synchronisation zu ersparen.

evtl.
beispielschema
erstellen für zwei
klassische transformationsklassen
und adapter vs.
unified transformation

Zitat einfügen?
Stefan Müller,
PCG? ;)

Die Aussage „Photorealistische Computergraphik ist die Simulation von Licht“ hat mich wohl auch inspiriert, den Simulationsbegriff allgemeiner aufzufassen und das Begriffspaar "Rendering und Physiksimulation" zu hinterfragen⁴.

Es sei bemerkt, dass weder eine Hypothese bestätigt noch widerlegt werden sollte, geschweige denn überhaupt eine (mir bekannte) Hypothese im Vorfeld existierte; Es sprechen etliche Argumente für eine Vereinheitlichung der Konzepte (geringerer Overhead durch Wegfall der Adapterklassen, evtl. Speicherverbrauch durch z.T. gemeinsame Nutzbarkeit von Daten), aber auch einige dagegen (Komplexität eines Systems, Anzahl an theoretischen

²<http://bulletphysics.org>

³<http://www.ogre3d.org>

⁴Auch wenn dieses Framework nicht vornehmlich auf physikalisch basierte, also photorealistische Beleuchtung ausgelegt ist, soll diese aufgrund des generischen Konzepts jedoch integrierbar sein.

Kombinationsmöglichkeiten steigt, viele sind unsinnig und müssen implizit oder explizit ausgeschlossen werden).

Für mich persönlich bringt die Bearbeitung dieser Fragestellung zahlreiche Vorteile; Ich muss ein wenig ausholen:

Schon als Kind war ich begeistert von technischen Geräten, auf denen interaktive Computergraphik möglich war; Sie sprechen sowohl das ästhetische Empfinden an, als auch bieten sie eine immer mächtigere Ergänzung- und Erweiterungsmöglichkeit zu unserer Realität an; Letztendlich stellten diese Geräte für mich wohl auch immer ein Symbol dafür dar, in wie weit die Menschheit inzwischen fähig ist, den Mikrokosmos zu verstehen und zu nutzen, damit demonstriert, dass sie zumindest die rezeptiven und motorischen Beschränkungen seiner Physiologie überwunden hat.

Die Freude an Schönheit und Technologie findet für mich in der Computergraphik und der sie ermöglichenen Hardware eine Verbindungsmöglichkeit; Die informatische Seite mit seinen Algorithmen als auch die technische Seite mit seinen Schaltungen faszinieren mich gleichermaßen; Auch das „große Ganze“ der Realisierung solcher Computergraphischen Systeme, das Engine-Design mit seinen softwaretechnischen Aspekten, interessiert mich. Ferner wollte ich schon immer „die Welt verstehen“, sowohl auf physikalisch-naturwissenschaftlich-technischer, als auch - aufgrund der system-immanenten Beschränkungen unseres Universums - auf metaphysischer Ebene⁵;

Und hier schließt sich der Kreis: Sowohl in der Philosophie als auch in der Informatik spielt das Konzept der Abstraktion eine wichtige Rolle; Nichts anderes tut eine „Unified Engine“: sie abstrahiert bestehende Konzepte teilgebiets-spezifischer Engines, wie z.B. Graphik und Physik; Ich erhoffe mir, dass mit dieser Abstraktion man in seinem konzeptionellen Denken der realen Welt ein Stück weit näher kommt; Die verfügbaren Rechenressourcen steigen, die Komplexität von Simulationen ebenfalls; Ob eine semantische Generalisierung von seit Jahrzehnten verwendeten Begriffen wie „Rendering“ und „Physiksimulation“, welche dieser Entwicklung angemessen sein soll, eher hilfreich oder verwirrend ist, kann eine weitere Interessante Frage sein, die ich jedoch nicht weiter empirisch untersucht habe.

Letztendlich verbindet dieses Thema also viele meiner Interessen, welche die gesamte Pipeline eines Virtual-Reality-Systems, vom Konzept einer Engine bis hin zu den Transistoren einer Graphikkarte, auf sämtlichen Abstraktionsstufen betreffen: Es gab mir die Möglichkeit,

- den Mehrwert einer Abstraktion gängiger Konzepte von Computergraphik und Physiksimulation zu erforschen

⁵ob der Begriff „Verständnis“ im letzten Falle ganz treffend ist, bleibt Ermessens-Sache

- die Erfahrungen im Engine-Design zu vertiefen
- die Erfahrungen im (graphischen) Echtzeit-Rendering zu vertiefen
- mich mit Physiksimulation (genauer: Simulation von Mechanik) zu beschäftigen, konkret mit Fluidsimulation
- mich in OpenGL 3 und 4 einzuarbeiten, drastisch entschlackten Versionen der Graphik-API, deren gesäuberte Struktur die Graphikprogrammierung wesentlich generischer macht und somit die Abstraktion erleichtert
- mich in OpenCL einzuarbeiten, den ersten offenen Standard für GPGPU⁶
- mich intensiver mit Graphikkarten-Hardware, der zu Zeit komplexesten und leistungsfähigsten Consumer-Hardware zu beschäftigen, aus purem Interesse und um die OpenCL-Implementierung effizienter zu gestalten

Die vielseitigen didaktischen Aspekte hatten bei dieser Themenwahl also ein größeres Gewicht als der Forschungsaspekt, was dem Ziel einer Bachelorarbeit angemessen ist.

1.1.2 Fluidsimulation

Warum ich mich bei der exemplarischen Implementierung einer mechanischen Simulationsdomäne für eine partikelbasierte Fluidsimulation entschieden habe, hat viele Gründe:

Wohingegen Rigid Body-Simulation in aktuellen Virtual-Reality-Anwendungen wie Computerspielen schon eine recht große Verbreitung erreicht hat, sucht man eine komplexere physikalisch basierte Fluidsimulation, die über eine 2,5 D- Heightfield-Implementation hinausgehen, noch vergebens; Das Ziel, eine deratige Fluidsimulation in einen Anwendungskontext zu integrieren, der langfristig über den einer Demo hinausgehen soll, hat also einen leicht pionierhaften Beigeschmack.

Mir ging es um eine Simulation, welche die Option einer möglichst breiten Integration in die virtuelle Welt bietet; Wohingegen sich Grid-basierte Verfahren aufgrund Möglichkeit zur Visualisierung per Ray-Casting sehr gut zur Simulation von Gasen eignen, sind Partikel-basierte Verfahren eher für Liquide geeignet, da bei Grid-basierten Verfahren die Volumen-Erhaltung eines Liquids durch zu Instabilität und physikalischer Implausibilität neigenden Level-Set-Berechnungen sichergestellt werden muss. Liquide beein-

Muss ich das irgendwie belegen?
Ich wüsste nicht,
wie/wo ;(

referenz auf
Peschel, evtl weitere
eigenschaften
wie curl und
sicherstellung der
inkompressibilität

referenz auf
gpu-gems-gridbased-
zeug, evtl jos stam

⁶General Purpose Graphics Processing Unit- Computing, die Nutzung der auf mas- siver Parallelität beruhenden Rechenleistung von Graphikkarten in nicht explizit Graphik-relevanten Kontexten

flussen aufgrund ihrer Dichte Objekte ihrer Umgebung im Alltag mechanisch stärker als Gase; Aufgrund dieser erhöhten gegenseitigen Beeinflussung von Fluid und Umgebung bevorzugte ich das Verfahren, welches Fluide besser simuliert.

Die Partikel-Domäne bietet außerdem eine theoretisch unendlich große Simulationsdomäne, wohingegen in der Grid-Domäne der Simulationsbereich auf das Gebiet beschränkt ist, welches explizit durch Voxel repräsentiert ist. Ferner lassen sich relativ einfach auch Rigid Bodies durch einen partikelbasierten Simulator simulieren, indem eine Repräsentation der Geometrie des Rigid Bodies als Partikel gewählt wird;

referenz auf
thomas steil und
GPU gems

Bei grid-basierten Verfahren lässt sich die Simulations-Domäne z.B. als Sammlung von 3D-Texturen oder nach einem bestimmten Schema organisierten 2D-Texturen repräsentieren; Hiermit wird die Simulation auf der GPU mithilfe von Graphik-APIs wie OpenGL ohne weiteres möglich, und wurde auch schon erfolgreich implementiert. Mein Anliegen war jedoch, explizit die Features moderner Graphikhardware und sie nutzender GPGPU-APIs wie Nvidia CUDA, Microsoft's DirectCompute oder OpenCL zu verwenden; Die Partikel-Domäne stellt damit eine größere Abgrenzung zu gewohnten Workflows auf der GPU dar. Vor allem die *Scattered Writes*, die Graphik-Apis nicht oder nur sehr indirekt ermöglichen, und die von so vielen Algorithmen benötigt werden, sollten zum Einsatz kommen dürfen.

ref auf peschel

Die Fluidsimulation stellt einen relativ „seichten“ Einstieg in die Welt der GPU-basierten Echtzeit-Physiksimulation dar:

- Es gibt schon zahlreiche Arbeiten zur Fluidsimulation, welche erfolgreich den Spagat zwischen Echtzeitfähigkeit und physikalischer Plausibilität gemeistert haben (siehe Kapitel 4.2.1.2);
- Fluide sind für gewöhnlich ein homogenes Medium, daher eignet sich bei Partikel-Ansatz für die Suche nach Nachbarpartikeln die Beschleunigungsstruktur des Uniform Grid besonders gut, wohingegen sich für Simulation von Festkörpern eher komplexere Beschleunigungsstrukturen wie Oct-Trees, Bounding Volume-Hierarchies oder kD-Trees anbieten, da diese sich besser an die inhomogenen Strukturen, Ausmaßen und Verteilungen der Objekte anpassen. ; Letztere Strukturen lassen sich schwerer auf die GPU mappen, welche als Stream-Prozessor nicht optimal für komplexe Kontrollflüsse und Datenstrukturen geeignet ist.
- Die Partikel lassen sich direkt als OpenGL-Vertices per Point Rendering darstellen, was während der Entwicklungsphase eine einfache Visualisierungsmöglichkeit bietet;
- Die gemeinsame Nutzung von Geometrie sowohl zur mechanischen

‘indirekt’ erläutern? transform feedback buffer, scattering über gl_Position und ein-pixel point rendering, siehe DA von Sinje Thiedemann?

Wo zum Henker soll ich dazu nun wieder ne Referenz herauskramen? Ich habe leider keine Idee, außer zig allgemeine Papers und Übcher zu überfliegen Zeitproblem!... das ist der Nachteil, wenn man so viel recherchiert hat, zunächst mal fürs Verständnis und oder oder den ganz groben Überblick, und nicht für eine umfassende wissenschaftliche Dokumentation: All den Input, der nur als Hintergrundwissen dienen sollte, jetzt noch wieder

ref auf späteren
abschnitt oder
OpenCL spec

hier bin ich mir
nicht mal sicher,
ob das stimmt;
meine Erinnerung
ist nur sehr vage,
diese Fragestellung
wurde in
meiner Literatur
höchstens ganz
marginal behan-
delt; Was tun?
Aussage weg
lassen, mich tod-
suchen, oder diese
gewagte These
nach bestem Wis-
sen und Gewissen
stehen lassen?

Simulation als auch zur Visualisierung mit OpenGL ist hiermit er- möglicht. OpenCL ermöglicht diese gemeinsame Nutzung explizit über gemeinsame Buffer-Nutzung.

- Nicht zuletzt ist die Mathematik bei Partikel-Simulation einfacher: Die „eulersche Sicht“ auf die Simulationsdomäne beim Grid-Ansatz erfordert einen Advektionsterm, der dank der „lagrange’schen Sicht“ bei Partikeln wegfällt; Außerdem erfordern Partikelsysteme keine Berechnungen zur Sicherstellung der Inkompressibilität⁷ oder Bewahrung des Volumens.

⁷Das heißt nicht, dass die Inkompressibilität automatisch gewährleistet ist; Im Gegenteil hat man öfter mit „Flummi-artigem“ Verhalten des Partikel-Fluids zu tun, weil diese eben nicht ohne weiteres forcierbar ist;

2 Überblick

2.1 Vision

Die langfristige Vision, die *Flewnit* begleitet, ist die Entwicklung eines interaktiven Paddel-Spiels unter Verwendung dieser Unified Engine mit ausgefeilter Fluid-Mechanik und -Visualisierung, partikelbasierten Rigid Bodies und Dreiecks-Mesh als Repräsentation für statische Kollisions-Geometrie; Spiele, in der große Mengen Fluid, die komplexer simuliert sind als durch Height-Fields⁸ einen integrativen Bestandteil der Spielmechanik ausmachen, sind mir nicht bekannt;

Von Dreiecks-Geoemtrie erhoffe ich mir eine genauere Repräsentation zur Kollisionsbehandlung, bei gleichzeitiger Ersparnis vieler Partikel, die sonst z.T. große Oberflächen repäsentieren müssten; Ferner könnte die Dreiecksstruktur später zur Simulation nicht-partikelbasierter Rigid Bodies verwendet werden;

2.2 Paradigmen

Vor dem Entwurf eines komplexen Softwaresystems mit einigen Zügen, die in etablierten Systemen keine so große Bedeutung haben, hat es Sinn, sich einige Paradigmen zu überlegen, welchen das System nach Möglichkeit folgen soll, um eine gewisse Konsistenz zu gewährleisten:

- Es wurde beim Entwurf der Unified Engine für jede Simulationsdomäne eine möglichst ähnliche Struktur von Klassen und ihren Beziehungen zueinander angestrebt. Diese Ähnlichkeit spiegelt sich nach Möglichkeit in einer gemeinsamen (manchmal abstrakten) Oberklasse eines jeden Konzeptes wider, wie z.B.:
 - dem Simulations-Objekt als solchem
 - der Geometrie
 - dem Material
 - der Szenen-Repräsentation

Auf diese Weise soll eine maximale *Symmetrie* zwischen den Domänen hergestellt werden, so dass domänen-bedingte Spezial-Behandlung von Objekten und Workflows minimiert wird;

- Es sollte eine Art Pipeline-Architektur entstehen, wo bestimmte Pipeline-Stages bestimmte Simulations-(Zwischen)- Ergebnisse implementieren, und ggfs. anderen Stages diese zur Verfügung stellen. Jede Simulationsdomäne hat seine eigene Pipeline; Dennoch können Interdependenzen bestehen;

⁸s. Kapitel 4.2.1.2 für mehr Informationen zu Height-Field-basierter Fluidsimulation

Diesen Interdependenzen wird durch eine Konzept-spezifische Verwaltung durch verschiedene Singleton-Manager-Klassen genüge getan; Ein und dasselbe Objekt kann von verschiedenen Managern in unterschiedlichem Zusammenhang verwaltet werden; Mehr dazu in Kapitel 3;

- Es sollen langfristig so viele Features (Visualisierungstechniken und -effekte, Simulationstechniken) wie möglich miteinander kombinierbar sein, sofern die Kombination nicht unsinnig ist;
- Es soll so viel wie möglich auf der GPU berechnet werden, um die massive Parallelität auszunutzen, und um nicht durch Buffer-Transfers, die die Aufteilung von Algorithmen in CPU- und GPU- Code meist mit sich bringen, auf den Bandbreiten- und Latenz- Flaschenhals der PCI-Express-Schnittstelle zu stoßen;
- Es soll immer das Potential gewahrt werden, dass aus dem Framework — außerhalb des Rahmens dieser Bachelorarbeit — tatsächlich noch eine Art *Unified Engine* entstehen kann; Somit sind „schnelle Hacks“, also unsaubere Programmier-Weisen, die mit geringstem Programmieraufwand ein bestimmtes Feature implementieren, überall dort unbedingt zu vermeiden, wo sie die konsistente Gesamstruktur des Systems zu bedrohen scheinen.

2.3 Begriffe

Im Zuge der angestrebten Vereinheitlichung der verschiedenen Simulationsdomänen müssen wir auch einige Begriffe verallgemeinern, welche in ihrer jahrzentelangen Tradition in der Terminologie der Computergraphik eine spezifische Bedeutung erhalten haben; Zur besserern Einordnung stellt Abbildung 1 ein grobes Schema dar, welches die klassische Verwendung verschiedener Engines und die einer Unified Engine gegenüber stellt:

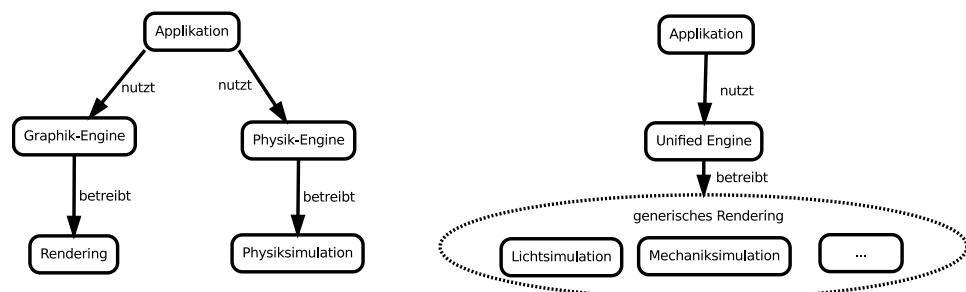


Abbildung 1: Gegenüberstellung von Verwendung und Begrifflichkeiten von klassischen Engines und einer Unified Engine

Rendering Im Wiktionary [?] wird das Verb *to render* u.a. umschrieben als:

„(transitive, computer graphics) To transform digital information in the form received from a repository into a display on a computer screen, or for other presentation to the user.“

Es geht also um die Transformation einer formalen Beschreibung in eine für einen menschlichen Benutzer wahrnehmbare Form. Diese muss entgegen der gewöhnlichen Verwendung des Begriffes nicht zwingend visuell, sondern kann z.B. auch akustischer oder haptischer Natur sein, übertragen durch Lautsprecher oder Force-Feedback-Devices.

Verallgemeinern wir den Begriff *Rendering* weiter, gemäß der Übersetzung der Verb-Form als *erbringen, machen* [?], und in Anlehnung an seine Ethymologie,

„From Old French *rendre* (“to render, to make”)“ [...] [?]

bietet sich eine freie Übersetzung als *Erzeugung eines Zustandes beliebiger Natur* an;

Unter diese generische (Um)-Deutung des Begriffes fällt nun auch *die Ausführung beliebig gearteter Simulation*.

Zu besseren Abgrenzung kann man von *generischen Rendering* und dem klassischen *visuellen Rendering* sprechen; Dies soll im weiteren Verlauf dieser Arbeit der Fall sein.

Eine *Unified Engine* (s.u.) betreibt also *generisches Rendering*.

Unified Engine Alternativ-Bezeichnung: *Unified Rendering Engine*;

Eine *Unified Engine* betreibt *generisches Rendering*, indem sie bestimmte Aspekte einer Welt⁹ simuliert. Darunter kann das klassische (visuelle) Rendering fallen, aber auch die Simulation von Geräuschen und von Mechanik, und beliebige weitere Domänen; Die Domänen sollen dabei durch Abstraktion gemeinsamer Eigenschaften so ähnlich wie möglich organisiert sein;

Simulation Das Begriffspaar *Rendering* und *Physiksimulation* ist im Kontext dieser versuchten Vereinheitlichung nicht mehr angemessen; Stattdessen sollten wir den Simulations-Gedanken aufgreifen und anstelle von *Rendering* lieber von *Licht-Simulation* sprechen; Auf diese Weise werden missverständliche abwechselnde Verwendung vom Begriff *Rendering* vermieden;

Der Begriff der *Physiksimulation* ist auch nicht ganz sauber, da streng genommen Licht auch ein physikalisches Phänomen ist, und somit

⁹Diese Welt muss dabei nicht zwingend unserer Realität ähneln oder entsprechen.

vom Begriff eingeschlossen wird, statt sich abzugrenzen; Es bietet sich die alternative und genauere Bezeichnung *Mechanik-Simulation* an; Die Quantenmechanik vor Augen (der Name spricht für sich) und damit den Umstand, dass auch Photonen an mechanischen Vorgängen teilnehmen, ist zwar selbst dies keine saubere Abgrenzung, aber auf dem angestrebten Niveau einer plausiblen (im Kontrast zur „korrekten“) Echtzeit-Simulation, welche mit der Newton'schen Physik auskommen wird, ist diese Abgrenzung klar genug.

GPU Computing Für gewöhnlich bekannt als Begriff, der GPGPU-Computing beschreibt in Abgrenzung zur Verwendung der GPU zur Berechnung von Bildern mittels Graphik-APIs, soll auch dieser Begriff im Folgenden einen Oberbegriff darstellen für beliebige Berechnungen, die auf der GPU ausgeführt werden, gleichgültig ob in einem GPGPU- oder einem Graphik-bezogenen Kontext; Die Abgrenzung der Domänen „Graphik“ und „General Purpose“ ist nicht zwingend anhand der benutzten APIs festzustellen; So kann man z.B. mit einer GPGPU-API Ray Tracing für visuelles Rendering betreiben, oder aber General Purpose-Berechnungen mit Graphik-APIs anstellen.¹⁰

Da im Computer jede Operation eine Form von Berechnung darstellt, erscheint die Verwendung von *GPU Computing* als Oberbegriff für legitim; GPU Computing unterteilt sich dann in *Graphik-Programmierung* (für gewöhnlich, aber nicht zwingend durch Verwendung von Graphik-APIs wie OpenGL oder Direct3D) und *GPGPU* (für gewöhnlich, aber nicht zwingend durch Verwendung von GPGPU-APIs wie CUDA, DirectCompute oder OpenCL).

Somit ist die erste Symmetrie zwischen den Simulationsdomänen durch eine Anpassung der Terminologie bewerkstelligt.

2.4 Schwerpunkte

Die Entwicklung eines solchen *Unified Frameworks*¹¹ umfasst sehr viele Aspekte, und einige werden wohl in dieser Ausarbeitung keine Erwähnung finden. Um die didaktischen Ziele von Seite 3 nicht aus den Augen zu verlieren, wurden folgende Schwerpunkte gesetzt:

¹⁰Vor der Zeit der *Compute Unified Device Architecture (CUDA)*-Devices und der gleichnamigen GPGPU-Programmier-Umgebung von Nvidia, welche mit dem G80-Chipsatz bzw. der GeForce 8800 GTX 2006 ihren Anfang nahm, war der „Missbrauch“ von Graphik-APIs für den Endbenutzer die einzige Möglichkeit, die massive Rechenleistung der Graphikkarte für generische Zwecke zu nutzen.

¹¹Von einer *Engine* möchte ich Kontext der Implementierung im Rahmen dieser Bachelorarbeit noch nicht sprechen, da dieser Begriff eine viel zu große Vollständigkeit der Implementation suggeriert.

Entwicklungs-Umgebung Nach Anfängen unter Windows 7 und Visual Studio 2010, gab es bald Probleme beim Compilen von Dependencies auf 64 Bit; Womöglich wäre es eine Frage der Geduld gewesen, jedoch habe ich dann zu Ubuntu Linux¹² und Eclipse¹³ in Kombination mit dem Cross-Platform Build-System CMake¹⁴ gewechselt. Mir war es sehr wichtig, ein System zu entwickeln, welches vollständig für die 64 Bit-Prozessor-Architektur compleit ist, da viele Register der heutigen 64Bit-Prozessoren sonst ungenutzt bleiben. Das Paket-Management der Linux-basierten Betriebssysteme und die konsequente Implementierung fast aller Programme in 64 Bit erleichtern das Einrichten der Dependencies ungemein; Schon alleine dafür hat sich der Umstieg gelohnt.¹⁵ Zum Schreiben der GPU-Programme habe ich *gedit*¹⁶ mit entsprechenden Syntax-Highlightings verwendet. Für die Source Code-Versionsverwaltung kam *git*¹⁷ zum Einsatz.

referenz finden,
bin mir da eher
unsicher

explizite Danksa-
gung?

Dependencies Das Endziel einer potenteren, modernen Engine sollte auf keinen Fall durch die Wahl suboptimaler Bibliotheken eingeschränkt werden; Andererseits sollten, um Compile- und Link-Zeiten gering zu halten und Konflikte zwischen Bibliotheken zu vermeiden, die Dependencies nicht zu komplex sein; Vor allem die Wahl des Fenster-Managers, der Input- Bibliothek und der Mathematik-Bibliothek musste deshalb mit Bedacht getroffen werden; Mehr dazu in Kapitel 3.1.

Nutzung moderner OpenGL-Features Mit OpenGL Version 3 erfuhr die offene Graphik-API eine gründliche Reinigung: Viele nicht mehr zeitgemäße Features wurden für deprecated erklärt und durch einige neue, meist generischere Features ersetzt; Das Resultat ist eine schlankere API mit mehr Verantwortung für den Programmierer über die Rendering-Pipeline; Man muss mehr „selbst machen“, hat aber auch mehr Macht. Diese Neuerung spielt dem Konzept einer Unified Engine geradezu in die Hände, da nun fast die gesamte vordefinierte Semantik wie `gl_FrontMaterial.shininess`, `gl_LightSource[i].diffuse`, `gl_NormalMatrix` oder `gl_MultiTexCoord2` entfernt wurde, und man jetzt fast alle benötigten Werte entweder als generische Vertex-Attributes übergibt oder selber explizit als Uniform - Variablen definiert und setzt; Diese neue „Freiheit der Semantik“ war eine große Inspiration, den Gedanken

¹²<http://www.ubuntu.com/>

¹³<http://www.eclipse.org/>

¹⁴<http://www.cmake.org/>

¹⁵Einen sehr sehr großen Dank möchte ich an dieser Stelle Lubosz Sarnecki aussprechen, der mich mit meiner zuvor sehr eingeschränkten Linux-Erfahrung unermüdlich mit Profi-Support bei der fortgeschrittenen Customization des Betriebssystems versorgt hat; Ohne ihn wäre mir dieser schnelle, weitgehend reibungslose Umstieg nicht gelungen.

¹⁶<http://projects.gnome.org/gedit/>

¹⁷<http://git-scm.com/>

einer Unified Engine zu fassen, wo es nun keinen Anlass mehr gab, OpenGL-State-Variablen implizit um zu interpretieren um bestimmte Effekte zu realisieren (Beispiel: Shadow-Map-Lookup-Matrix in eine der Textur-Matrizen laden).

Es sollte mindestens ein OpenGL Kontext im Core Profile der Version 3.3 benutzt werden; Das Core Profile stellt sicher, dass Routinen und Flags, die in der Spezifikation als deprecated gekennzeichnet sind, einen Fehler produzieren; Auf diese Weise wird die Programmierung mit nur der „modernen“ Untermenge der OpenGL-API forciert; Optional sollte ein OpenGL 4 Kontext erstellt werden können, sofern unterstützende Hardware existiert und dies vom Benutzer erwünscht ist; Wo es sinnvoll und angebracht war, sollten moderne Features von OpenGL verwendet werden; Mehr dazu in Kapitel 4.1.8.

Template-Engine Da GLSL¹⁸ und OpenCL C, die verwendeten Sprachen, mit denen die GPU programmiert werden kann, nicht objekt-orientiert sind, (zumindest OpenGL 3 und OpenCL 1.0) keine Mechanismen zum Überladen von Funktionen haben, und noch nicht einmal mal eine

`#include`-Direktive existiert, tendieren diese GPU-Programme, die sich einen erheblichen Anteil an ihrem Code untereinander teilen, zur Code-Vervielfachung in den einzelnen Quelldateien; Dies schränkt die Wartbarkeit und bequeme Änderbarkeit und zuweilen auch die Lesbarkeit enorm ein;

Abhilfe schafft hierbei die Nutzung einer String-Template-Engine namens *Grantlee*¹⁹; Mit ihrer Hilfe lassen sich zur Laufzeit in Abhängigkeit von aktuellen Parametern angepasste GPU-Programme generieren, die nur den aktuell nötigen Code enthalten; Somit sind die generierten Programme lesbarer als wenn man sie über klassische bedingte Komplilierung (mit `#ifdef FEATURE_XY ... #endif`-Direktiven) geschrieben hätte; Weitere Einzelheiten sind in ?? zu finden;

Implementierung und Kombination gängiger visueller Effekte Um dem softwaretechnischen Unterbau schließlich etwas Leben einzuhauen, wurden einige visuelle Effekte unter Nutzung von OpenGL3/4 implementiert; Dank der Template-Engine lassen sich alle Effekte – sofern sinnvoll – Zur Laufzeit in beliebiger Kombination hinzufügen oder abschalten. Die Effekte werden in Kapitel 4.1.10 detaillierter vorgestellt.

Buffer-Abstraktion Der zentrale Dreh- und Angelpunkt der Unified Engine ist der *Buffer*;

¹⁸OpenGL Shading Language

¹⁹www.grantlee.org/

Ist „Buffer“ für gewöhnlich die Bezeichnung eines allokierten Speicherbereiches zur Nutzung durch ein Programm, verkompliziert sich diese simple Sicht auf einen Buffer durch die GPU Computing- APIs, erstens, weil man zwischen Host- und Device- Memory unterscheiden muss, zweitens, weil die GPU Computing- APIs den Buffern verschiedene Semantiken zuschreiben (generischer Buffer, Vertex Attribute Buffer, Vertex Index Buffer, Uniform Buffer, Transform Feedback Buffer, Texture Buffer, Render Buffer, verschiedene Texturtypen etc.), die je nach API zwischen den einzelnen Verwendungs-Kontexten (Host, OpenGL, OpenCL) kompatibel zueinander sind oder eben auch nicht;²⁰ Die einzelnen Buffertypen haben trotz ihrer semantischen Unterschiede und verschiedensten zugehörigen API-Routinen einige konzeptionelle Gemeinsamkeiten:

Die meisten Buffertypen haben irgendeine Form von folgenden assoziierten Operationen:

- Allokation von Speicher
- Freigabe von Speicher
- Schreiben
- Lesen
- Kopieren
- Mappen von device-Memory zu Host-Memory
- Spezifikation von Semantik (insb. bei OpenGL Buffers)
- Spezifikation von internen Datentypen, ggfs. Channel-Layout etc.
- Synchronisieren vor dem nächsten Zugriff
- Acquirering für einen Kontext zur Nutzung von API- Interoperabilität

Ebenfalls ist eine Menge Meta-Information vielen Buffertypen gemeinsam:

- Name
- Buffergröße
- Buffertyp
- Buffer-Semantik
- interne Datentypen

²⁰Mit CUDA ist es z.B ohne weiteres möglich, einen beliebigen GPU-Buffer an eine Textur-Einheit zu binden und entsprechend wie eine Textur zu samplen, und umgekehrt; OpenCL erlaubt dies nicht; Hier muss man sich im Vorfeld entscheiden, ob man einen „normalen“ Buffer oder eine Textur haben will.

- weitere Buffertyp-spezifische Meta-Informationen
- Information der beteiligten Kontexte (z.B. nur Host-Memory, reiner OpenGL-Buffer, CL/GL-interop-Buffer mit oder ohne assoziierten Host memory etc.)

Je nach Buffertyp und assoziierter API unterscheiden sich die Routinen, um diese Operationen auszuführen bzw. diese Meta-Informationen auszulesen, teilweise erheblich; Um dem Benutzer der Unified Engine die Bürde der API-spezifischen Operationen abzunehmen, und möglichst viel Meta-Information geschlossen zur Verfügung zu stellen, bietet sich an, ein vereinheitlichtes Interface bereit zu stellen, von dem konkrete Buffertypen abgeleitet werden können, die die entsprechenden Operationen implementieren; Hieraus entstanden eine Reihe von Buffer-Und Texturklassen, welche zugehörige Meta-Informationen enthalten. Eine detaillierte Beschreibung findet sich in Abschnitt 3.12.

Effiziente Verwendung von OpenCL Bei der OpenCL-Implementierung lag der Schwerpunkt sowohl bei der algorithmischen Effizienz als auch bei hardwarespezifischen Optimierungen; Es werden Parameter wie z.B. die Größe des lokalen Speichers einer OpenCL Compute Unit abgefragt, und in Verbund mit den benutzerdefinierten Simulationsparametern die Konstanten im OpenCL-Code mit der Template Engine und Workload-Parameter und Buffergrößen auf Seiten der Applikation entsprechend gesetzt; Für Details sei auf 4.2.1.6 verwiesen;

3 Systemarchitektur

Dieses Kapitel soll ein Gefühl für die Komponenten von *Flewnit* und ihre Zusammenhänge vermitteln. Die Komponenten „in Aktion“ werden im Detail im Verlauf des Kapitels 4 beschrieben. Dieses Kapitel soll die groben Konzepte und Ideen teilweise anhand von Beispielen vorstellen. Nicht jedes Feature, welches erwähnt wird, ist auch tatsächlich vollständig implementiert und getestet. Über den Status der Implementierung zum Zeitpunkt der Abgabe dieser Ausarbeitung gebe ich am Ende dieses Kapitels (Abschnitt 3.15) einen Überblick.

Das System wurde in C++ als (wahlweise statisch oder dynamisch zu linkende) Bibliothek implementiert. Der Code der GPU-Programme ist in GLSL bzw. OpenCL C verfasst.

3.1 Dependencies

Zunächst sollen die verwendeten Third-Party-Bibliotheken kurz vorgestellt werden:

OpenGL3/4 Die schon mehrfach erwähnten modernen Versionen der *Open Graphics Library*, der offenen API der Khronos Group zur hardwarebeschleunigten Graphik-Programmierung auf Basis der Dreieck-Rasterisierung. Um die Programmierung ohne Legacy-Routinen nicht erst zur Laufzeit über einen OpenCL-Error durch Verwendung eines Core-Profiles zu erzwingen, gibt es einen OpenGL-Header namens „gl3.h“²¹, der in Kombination mit der entsprechenden Präprozessor-Definition `#define GL3_PROTOTYPES 1` schon zur Compile-Zeit nur die non-deprecated Routinen zur Verfügung stellt.

evtl treffenderen Ausdruck finden:
Scanline-basiert
oder was auch immer

OpenCL 1.0 Die *Open Computing Language*, erste Version der noch jungen API für massiv parallele Programmierung²², wie OpenGL von der Khronos Group verwaltet; sie stellt den ersten offenen Standard für GPGPU dar, d.h., die Verwendung der API ist nicht mehr an eine bestimmte Hardware (wie bei Nvidia CUDA) oder ein bestimmtes Betriebssystem (wie Microsofts DirectCompute) gebunden.

Zur Zeit der Implementierung waren noch keine Non-Developer-Treiber für OpenCL 1.1 verfügbar, außerdem gab es kein Feature dieser Version, welches ich dringend benötigt hätte. Deshalb habe ich die Version 1.0 verwendet.

Es gibt einen C++-Wrapper der C-API, welcher stark auf C++-Templates basiert und in einer einzigen Headerdatei implementiert ist. Dieser

²¹beziehbar unter <http://www.opengl.org/registry/>

²²die GPGPU-Computing einschließt

ist direkt von der Khronos-Homepage²³ beziehbar. Diesen Wrapper habe ich verwendet, da er die Nutzung der API wesentlich eleganter macht.

GLFW 2.7 Wie auf Seite 11 angedeutet, waren mir folgende Dinge wichtig, damit die Einsetzbarkeit des Frameworks in professionelleren Kontexten nicht schon im Vorfeld verbaut ist:

- Option auf Fullscreen
- Option auf Multisampling
- Die Möglichkeit der Erstellung eines OpenGL-Kontextes einer frei wählbaren Version mit Option zwischen Core- und Compatibility-Profile
- Option auf „Mouse Grab“, so dass man wie in einem Computerspiel mit ausgeblendetem Mauszeiger nur durch Bewegung der Maus ohne Bildschirm-/Fenster-Grenzen die virtuelle Kamera rotieren kann;
- „Input events“, d.h. Aktualisierungen von Benutzereingaben sollen häufig und mit minimaler Latenz geschehen, außerdem so unabhängig wie möglich von der Framerate sein; Nach Möglichkeit sollten Input-Updates zumindest aktiv abfragbar sein (im Gegensatz zum passiven Warten darauf, dass von der Input-Library eine Callback-Funktion aufgerufen wird)
- Es soll volle Kontrolle über die "Render-Loop" geben, so dass man nicht den Kontrollfluss an eine Funktion übergibt, die womöglich nie zurückkehrt und weiteren Kontrollfluss durch das Benutzerprogramm nur über Callback-Funktionen ermöglicht (wie `glutEnterMainLoop()` beim in die Jahre gekommenen *GLUT*). Ein derartiges Konstrukt ist einer Engine nicht würdig und verhindert womöglich sauberes Herunterfahren und Neu-Initialisierung, wie es z.B. beim Wechseln einer Szene oder eines fundamentalen globalen Settings nötig sein könnte.

GLFW²⁴ in der Version 2.7, die zum Zeitpunkt der Implementation aktuellste stabile Version, erfüllt diese Forderungen, und findet damit in *Flewinit* Einsatz sowohl im Fenster- als auch im Input-Manager. Die Timing-Funktionalität wird ebenfalls von GLFW übernommen.

OpenGL Mathematics (GLM) Seit sämtlicher Mathematik-bezogener OpenGL-State inklusive zugehöriger built-in-Variablen und -Funktionen wie `gl_ModelViewProjectionMatrix` oder `ftransform()` in GLSL abgeschafft wurden, führt um eine Bibliothek für Vektor- und Matrix-Algebra kein

²³<http://www.khronos.org/registry/cl/>

²⁴<http://www.glfw.org/>

Weg mehr herum. Ich habe mich für *GLM*²⁵ entschieden, da sie klein und dennoch mächtig ist, und einige Convenience-Functions hat; QT hat ebenfalls eine Mathe-Bibliothek, verwendet jedoch `double`, also 64bit-Fließkomma-Werte als Basis-Datentyp, was das direkte Übergeben als Array von `float`-Uniforms an die OpenGL-Shader verhindert. Zwar unterstützen OpenGL und moderne Graphikkarten `double` nativ, jedoch mit drastisch geringerer Geschwindigkeit, da weniger Recheneinheiten für diesen Datentypen zur Verfügung stehen.
 Es sei bemerkt, dass ich persönlich die direkte Verwendung einer C++ - Mathe-Bibliothek mit überladenen Operatoren und eigener Akkumulation von Matrizen wesentlich angenehmer und eleganter finde als das zähe Hantieren mit der C-API zum modifizieren des OpenGL-State mit seinen Matrix-Modes.

Grantlee Die bereits erwähnte Template-Engine; die Syntax entspricht der der Template-Sprache des *Django* web Frameworks²⁶.

Es lassen sich durch die Applikation Werte an die Template-Engine übergeben, anhand derer dann die Code-Generierung Kontrolliert wird, bzw welche durch Einschluss in doppelte geschwungene Klammern direkt eingefügt werden können.

Beispiel:

```

1 vec4 fragmentColor =
2 {%
3     if SHADING_FEATURE_DIFFUSE_TEXTURED %
4         texture(decalTexture, input.texCoords.xy);
5     {%
6     else %
7     color;
8     {%
9     endif %
10 }
```

```

1 #define NUM_BITS_PER_KEY_TO_SORT ( {{numBitsPerKey}} )
```

Die Engine stellt einen Vererbungs-Mechanismus bereit:
 Auf diese Weise kann z.B. die Datei „particleSimulationTemplate.cl“ verschiedene „Code-Blocks“ definieren wie z.B.

```

1 {%
2     block performSPHCalculations %
3         {%
4             comment %
5                 the core of the physics simulation:
6                 accumulate all relevant values
7                 (density, pressure force, viscosity force etc ...)
8             {%
9             endcomment %
10 }
```

²⁵<http://glm.g-truc.net/>

²⁶<https://www.djangoproject.com/>

, die von anderen Dateien geerbt und entsprechend angepasst implementiert werden.

„updateDensity.cl“ erbt von dieser Datei durch die Directive `{% extends "particleSimulationTemplate.cl" %}` und implementiert die Dichte-Berechnungen, ohne dass der umschließende (nicht unerheblich lange und komplexe) Kontrollfluss-Code wiederholt werden muss:

```
1  {% block performSPHCalculations %}
2    if( BELONGS_TO_FLUID(
3      GET_CURRENT_NEIGHBOUR_PARTICLE_OBJECT_ID   ) )
4    {
5      ownDensity +=
6        cObjectGenericFeatures [ ←
7          GET_CURRENT_NEIGHBOUR_PARTICLE_OBJECT_ID  ].←
8          massPerParticle
9        * poly6( ownPosition -  GET_CURRENT_NEIGHBOUR_POS , ←
10           cSimParams );
11    }
12  {% endblock performSPHCalculations %}
```

Assimp Die *Open Asset Import Library*²⁷ ermöglicht das Auslesen von Szenen direkt aus .blend-Dateien, dem nativen Datenformat des exzellenten Free Software – 3D- Modellierungsprogramms *Blender*²⁸. Somit entfällt der Umständliche Export in ein Zwischenformat.

TinyXML Um zu gewährleisten, dass das System schon in der frühen Entwicklungsphase weitgehend ohne Recompile-benötigenden „Hard-Codes“ konfigurierbar ist, wurde TinyXML verwendet, um eine XML-config-Datei zu parsen.

Außerdem haben manche Komponenten der *Boost-Libraries*²⁹ verwendet.

3.2 Klassendiagramm

Abbildung 2 und 3 zeigen ein vereinfachtes Klassendiagramm von *Flewnit*. Die roten Klassen stellen wichtige Meta-Informationen dar, um Kontrollfluss und Objekt-Erstellung zu delegieren. In den folgenden Unterabschnitten werden die einzelnen Komponenten detaillierter vorgestellt.

3.3 BasicObject und Memory Tracking

Die Bibliothek soll zur Laufzeit kontrolliert herunterfahr- und re-initialisierbar sein, um eine vielseitige und flexible Anwendung zu gewährleisten.

²⁷<http://assimp.sourceforge.net/>

²⁸<http://www.blender.org/>

²⁹<http://www.boost.org/>

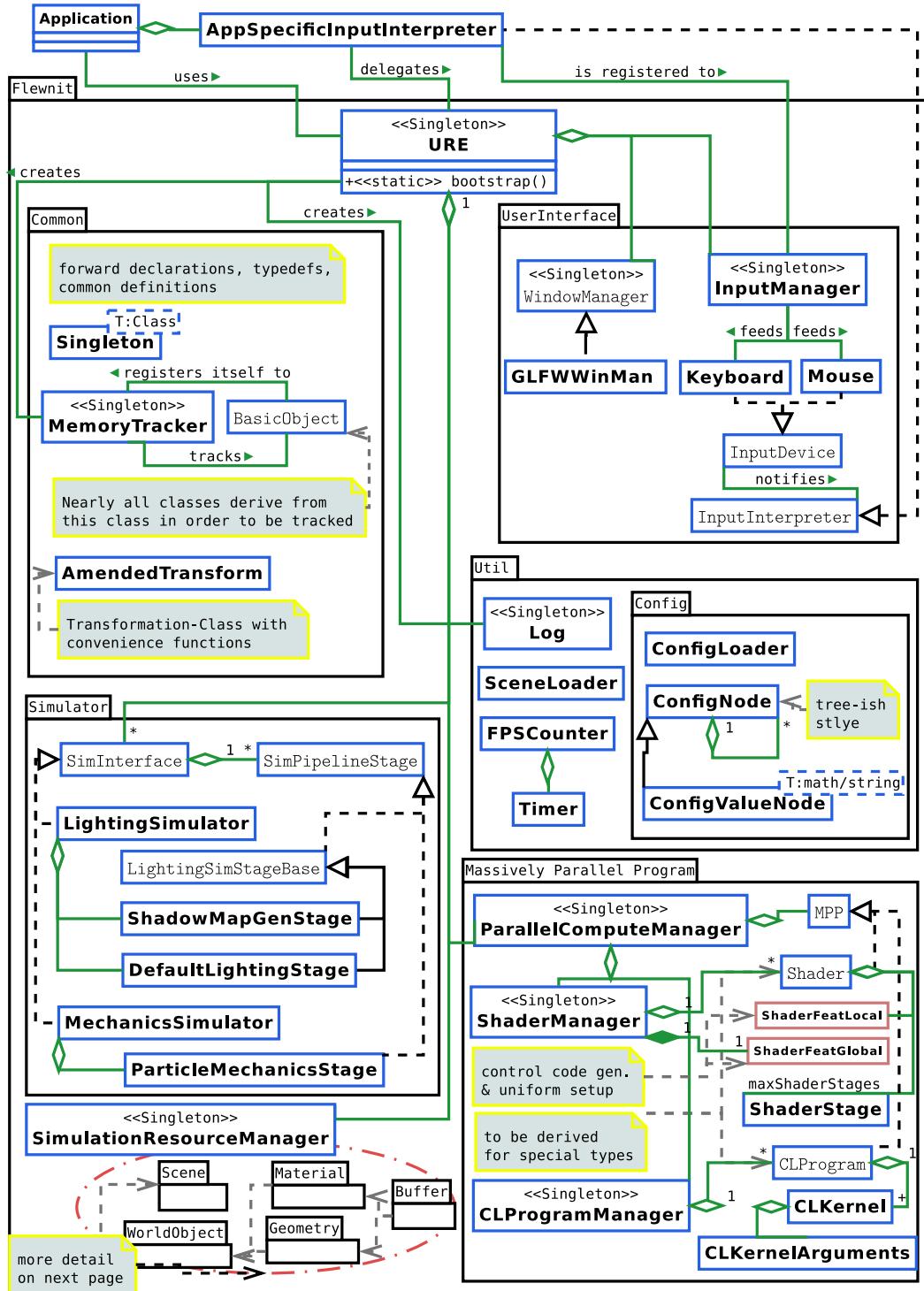


Abbildung 2: Klassendiagramm des Gesamtsystems, Teil 1

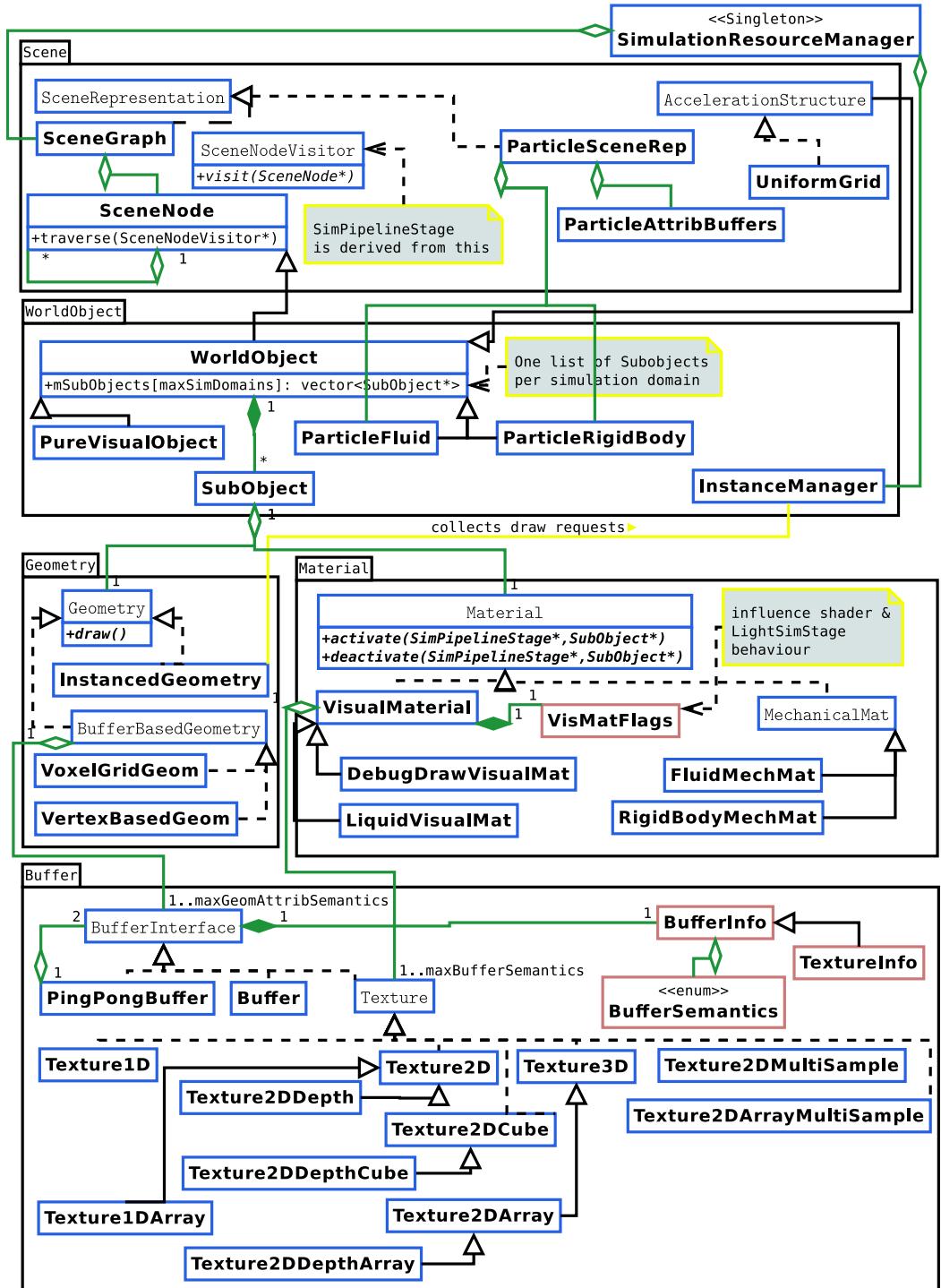


Abbildung 3: Klassendiagramm des Gesamtsystems, Teil 2

Da bei C++ das Speicher-Management dem Programmierer überlassen ist, und man daher vor allem bei massiver Nutzung von Pointern schnell den Überblick verliert,

- welches Objekt welche anderen erschaffen hat
- welche Objekte eine Klasse „besitzt“ und somit für ihre Speicher-Freigabe zuständig ist bzw
- auf welche Objekte eine Klasse nur ein Handle für Verwaltung oder Backtracking hat, ohne für Erschaffung oder Löschung zuständig zu sein

wurde in Anbetracht der oberen Forderung ein sehr simples Meta-Object-System realisiert, welches Klassen-Namen und Speicherverbrauch eines Objektes feststellt sowie jedem Objekt eine *unique ID* zuweist. Zweck dieses Meta Object System-Systems war es, Speicherlecks durch Nicht-Löschen von Objekten während der Entwicklung zu finden;

Dafür erbt beinahe jede Klasse in *Flewnit* von *BasicObject*. Diese Klasse registriert sich automatisch bei der *MemoryTracker*-Singleton-Instanz, bekommt dort eine ID zugewiesen. Diese ID hat noch keine Anwendung, könnte sich aber in einem Netzwerk-Kontext als nützlich erweisen.

Wie bei Qt das *Q_OBJECT*-Makro muss in jede von *BasicObject* erbende Klasse das Makro *FLEWNIT_BASIC_OBJECT_DECLARATIONS* eingetragen werden; Dieses Makro definiert die Implementatoin einer virtuellen Funktion, welche die Meta-Information setzen:

```
1 #define FLEWNIT_BASIC_OBJECT DECLARATIONS \
2     public:\\
3         virtual void initBasicObject() \\
4     { \
5         mMemoryFootPrint = (int) sizeof(*this); \
6         mClassName = String(typeid(*this).name()); \
7     } \
8     private:
```

Diese umständliche Implementierung hat die Ursache, dass Typ-Informationen der Blatt-Klasse einer Klassenhierarchie erst dann zur Verfügung stehen, wenn alle Konstruktoren der Basis-Klassen zurückgekehrt sind, weshalb der obere Code also nicht im Konstruktor der Basisklasse stehen darf, da der *this*-Pointer noch nicht die Informationen der Blatt-Klasse enthält. Ferner, um dem Programmierer nicht zuzumuten, *initBasicObject()* in jedem Konstruktor aufzurufen (das führt bei tiefen Klassenhierarchien zu unnötigen ausführungen dieser Funktion, da nur die Info der Blatt-Klasse benötigt wird, außerdem entstünde so eine weitere Vergesslichkeits-Fehlerquelle, die durch das Meta Object System ja gerade *vermieden* werden soll), stellt der Memory Tracker eine Routine *updateMemoryTrackingInfo()* bereit, die aufgerufen

werden sollte, wenn man auf valide Meta-Information zugreifen will. Aufpassen muss man bei tiefen Klassenhierarchien, da hier der Compiler keinen Fehler erzeugt, wenn eine Blatt-Klasse das Makro nicht eingetragen hat, da eine Implementierung der Funktion ja bereits durch eine Oberklasse existiert.

Über bedingte Kompilierung lässt sich diese Meta-Funktionalität deaktivieren. Ein „richtiges“ Meta Object System wie das von Qt wollte ich nicht verwenden, da es für meine Zwecke zu komplex und mächtig ist, und ich nicht mit Kanonen auf Spatzen schießen wollte. Ohnehin gibt es Tools wie Valgrind, mit denen man Speicherlecks finden kann.

Letztendlich ist meine Lösung nicht wirklich elegant (auch wenn ich ohne Meta-Object-Compiler (MOC) keine bessere Lösung gefunden habe) und daher eher als eine Spielerei anzusehen mit dem Zwecke, die C++-Interna zu verinnerlichen und zu jeder Zeit daran erinnert zu werden, Destruktoren zu implementieren und sich über „Besitz- und Verwaltungs-Verhältnisse“ der Klassen Gedanken zu machen.

Der `MemoryTracker` verfolgt auch sämtliche Allokationen der `BufferInterface`-Klasse, also der Basisklasse sämtlicher Buffer-Typen;

3.4 AmendedTransform

Im Zuge der Überlegungen zum Szenenraphen, der Kamera und des visuellen Renderings vieler Objekte per Hardware-Instancing entschied ich mich, die klassische 4x4-Transformationsmatrix zu wrappen und mit convenience functions anzureichern, siehe Listing 1.

Der Zweck dieser Klasse ist es, Transformationsmatrizen anhand „anschaulicher“ Parameter (Position, Richtung, Up-Vektor, Skalierung) zu definieren, und diese Parameter auch nach Akkumulationen, Modifikationen und/oder Animationen zur Verfügung zu haben.

Der Fokus liegt hier klar auf der Bequemlichkeit für den Entwickler, nicht auf der Performanz. Sobald sehr große, dynamisch animierte und tiefe Szenenraphen zum Einsatz kommen, könnte diese Klasse zum Flaschenhals werden. Dann ist vielleicht ein Refactoring vonnöten; Vorerst scheint mir die Implementation jedoch angemessen.

So lässt sich z.B. bequem ein `Camera`-Objekt, welches selber von `WorldObject` erbt, welches wiederum von `SceneNode` abgeleitet ist, an eine andere `SceneNode` anhängen, und aus der automatisch berechneten globalen Transformation erhält man durch `AmendedTransform::getLookAtMatrix()` direkt die lookAt-Matrix, so dass die Camera-Klasse selber sich nur noch um seine Projektionsmatrix kümmern muss. Auch typische Animationen werden von der Klasse übernommen. Ob dies guter Stil ist, oder doch lieber in die `SceneNode`-Klasse ausgelagert werden sollte, ist eine Frage, die ich noch nicht abschließend beantworten kann. Es wurde hier dem Paradigma der Datenlokalität gefol-

gt, evtl. auf Kosten der Separation of Concerns. Eine weitere Erörterung lasse ich aus, da das Refactoring in diesem Falle nicht zu aufwändig wäre;

Es sei bemerkt, dass durch die Menge an Matrizen, die an den Vertex Shader beim Instanced Rendering übergeben werden müssen, wenn man für alle Fälle gewappnet sein will³⁰, nicht unerheblich ist, da für jede Instanz ein anderes Matrix-Set übergeben werden muss; Die Menge an Daten, die per Uniform-Variable übergeben werden kann, ist begrenzt, daher hängt die maximale Anzahl an Instanzen, die pro Draw-Call gezeichnet werden können, direkt davon ab, wie viele Daten pro Instanz übergeben werden müssen;

Dieser Umstand hat mich dazu verleitet, die Normal Matrix, die transponierte Inverse der ModelView-Matrix, „einzusparen“: da Normalen und Tangenten durch Interpolation für den Fragment Shader ohnehin normalisiert werden müssen, spielt beim Ergebnis der Transformation eines Vektors³¹ nur die Richtung eine Rolle, nicht der Betrag. Die Richtung ändert sich nicht, wenn eine uniforme Skalierung in der Matrix steckt, wenn also der Skalierungsfaktor pro Raumdimension überall gleich ist. Der Translations-Anteil einer Matrix spielt bei Vektoren keine Rolle. Aus diesen Umständen lässt sich folgern, dass man zur Transformation von Vektoren ebenfalls die Model- bzw ModelView- Matrix verwenden kann, und *keine* dedizierte Normal Matrix braucht, solange nur Translation, Rotation und uniforme Skalierung in der Matrix akkumuliert sind.

Eben dies versuche ich, durch die `AmendedTransform`-Klasse zu forcieren, indem nur ein eingeschränkter `public`-Konstruktor existiert, direkte Initialisierung per Matrix-Konstruktor nur für bestimmte `friend`-Klassen erlaubt sind, und auch dann diese Matrix validiert wird. Die Einschränkung erscheint mir legitim, da eine Uniform Rendering Engine den Anspruch hat, zumindest im Ansatz „physikalisch basiert“ zu sein; Deshalb sollten sich Skalierungen ohnehin nur selten zur Laufzeit ändern, und dann nur einheitlich. So kann man im Vorfeld die Geometrie so anpassen, dass keine nicht-uniforme Skalierung in der Transformationsmatrix auftaucht.

Eine `AmendedTransform` ist insofern mit der `SceneNode`-Klasse verzahnt, dass falls eine `AmendedTransform`-Instanz als Transformation einer Scene-Node dient, erstere ein Handle auf die Scene Node bekommt. Auf diese Weise kann die Scene Node immer automatisch informiert werden, wenn sich bei der `AmendedTransform` etwas geändert hat (es müssen dann ggfs. die Kinder der Scene-Node aktualisiert werden, z.B. für eventuelles Update von Bounding Boxes). Auf diese Weise muss weder die `AmendedTransform` nochmal durch die `SceneNode` gewrappt werden, noch können Seiteneffekte durch direk-

³⁰z.B. für Layered Rendering in verschiedene Texturen, z.B. in ein Textur-Array zur Generierung mehrerer Spotlight-Shadowmaps mit einem Draw-Call; Hierbei braucht es explizit nur die Model-Matrix, da es pro Layer verschiedene View-Matrizes (pro Spotlight eine) gibt, von daher nicht im Vorfeld eine ModelView Matrix akkumuliert werden kann

³¹mit homogener Koordinate 0, im Gegensatz zu Punkten, homog. Koord. 1

te Manipulation der Transformation auftreten, sollte der Programmierer vergessen, anschließend die Scenenode über ihre Transformations-Änderung zu informieren. Die Scene-Node-Kopplung ist rein intern, und braucht den Benutzer der Klasse nicht weiter zu interessieren, außer in der Hinsicht, dass Scene-Node-Manipulation direkt geschieht und die `SceneNode`-Klasse kein eigenes Interface zur manipulation von Tranformationen bereit stellt.

Listing 1: AmendedTransform Klassendefinition, gekürzt

```

1  class AmendedTransform
2  {
3  public:
4      AmendedTransform(
5          const Vector3D& position = Vector3D(0.0f,0.0f,0.0f),
6          // (0,0,-1) is assumed as initial orientation, express any deviation ←
7          // in euler angles(radians)
8          const Vector3D& direction = Vector3D(0.0f,0.0f,-1.0f),
9          const Vector3D& upVector = Vector3D(0.0f,1.0f,0.0f),
10         float scale = 1.0f);
11     AmendedTransform(const AmendedTransform& rhs);
12     virtual ~AmendedTransform();
13
14     AmendedTransform operator*(const AmendedTransform& rhs) const;
15     // the assignment operators copy only the transformation values, not the ←
16     // SceneNode-Related info
17     const AmendedTransform& operator*=(const AmendedTransform& rhs);
18     const AmendedTransform& operator=(const AmendedTransform& rhs);
19
20     // accum: translationMatrix * rotationMatrix * scaleMatrix;
21     const Matrix4x4& getTotalTransform() const;
22     // convenience functions:
23     // unscaled, i.e. orthonormal rotation matrix:
24     Matrix3x3 getRotationMatrix() const;
25     // inverse of (translationMat*rotationMat);
26     Matrix4x4 getLookAtMatrix() const;
27     Matrix4x4 getScaleMatrix() const;
28     Matrix4x4 getInverseScaleMatrix() const;
29     static bool matricesAreEqual(const Matrix4x4& lhs, const Matrix4x4& rhs ←
30         );
31     AmendedTransform getInverse() const;
32
33     /*... getter and setter for pos, dir, up omitted */
34
35     // "animation" functions:
36     void moveRelativeToDirection(float forwardBackward, float rightLeft, ←
37         float upDown);
38     // change direction by rotating it angleDegrees degrees around cross(←
39     // direction, upVector)
40     void pitchRelativeToDirection(float angleDegrees);
41     // change direction by rotating it angleDegrees degrees around the ←
42     // upVector;
43     void yawRelativeToUpVector(float angleDegrees);
44
45 protected:
46     // backtrace pointer to tell the scene node to update itself after its ←
47     // transform has been modified directly;
48     // mOwningScenNode->transformChanged(bool global) will be called by any ←
49     // setter function of this class if there is an
50     // associated SceneNode;

```

```

43 SceneNode* mOwningSceneNode;
44 //flag needed by Scene node to update itself appropriately
45 bool mIsGlobalTransform;
46
47 friend class SceneNode;
48 void setOwningSceneNode(SceneNode*node, bool isGlobalTransform)
49 {mOwningSceneNode = node; mIsGlobalTransform=isGlobalTransform;}
50
51
52 Vector3D mPosition;
53 Vector3D mDirection;
54 Vector3D mUpVector;
55 float mScale;
56
57 Matrix4x4 mAccumTranslationRotationScaleMatrix;
58
59 //construct the transformation matrix from current pos,dir,up,scale; is←
60 //called by any constructor and related setter;
61 //validates members, if possible
62 void setup();
63
64 friend class Loader;//the loader may set transformation matrices, as ←
65 //they are read directly from Scene descriptions ;
66 //protected matrix-constructor to omit that the user passes a non←
67 //conforming matrix
68 AmendedTransform(const Matrix4x4& transform);
69
70 };

```

3.5 Die *Unified Rendering Engine* (URE)

Dies ist die Singleton-Klasse, die sämtliche Komponenten besitzt/verwaltet, und über welche direkt oder indirekt alle Daten beschafft oder Funktionalität aufgerufen werden kann, die nach außen hin verfügbar sein soll.

3.6 Das UserInterface-Paket

Das User Interface hat einen abstrakten Window-Manager-Interface, welches von einer Implementation, die *GLFW* verwendet, realisiert wird; Aus der geprästen Config wird die gewünschte OpenGL-Kontext-Version, Auflösung, Multisamples-Count, Fullscreen-Option etc. ausgelesen und ein entsprechendes Fenster mit OpenGL-Kontext erzeugt.

Die Input-Verwaltung ist im Gegensatz zum Fenstermanager nicht vollständig von der verwendeten Input-Bibliothek (ebenfalls *GLFW*) abstrahiert, da beim Parsen des Inputs die von *GLFW* definierten Makros direkt verwendet werden, wie z.B `GLFW_KEY_ENTER`. Da die Austauschbarkeit von verwendeten Bibliotheken zwar erwünscht ist, aber ich mich nicht verzetteln wollte, bleibt die Abstraktion des Inputs noch ein langfristiges „TODO“ mit geringerer Priorität.

Eine GUI wurde leider noch nicht implementiert, vor allem aus dem Grund, dass vorhandene GUI-Toolkits wie z.B. *AntTweakBar*³² Legacy-OpenGL-Routinen verwenden, was die Verwendung eines modernen Core-Profiles ausschließen würde. Da langfristig ein konsistentes Erscheinungsbild angestrebt wird, welches seine GUI vollständig in das OpenGL-Rendering-Fenster integriert hat, kam auch die Verwendung von GUI-Toolkits wie dem von Qt nicht in Frage.

Letztendlich wollte ich schon immer mal eine kleine GUI entwickeln, wozu ich jedoch erwartungsgemäß im Rahmen dieser Arbeit keine Zeit hatte. Die GUI verbleibt also als ein „TODO“ von mittlerer Priorität.

3.7 Das Simulator-Paket

Wie bereits erwähnt, soll jede Simulationsdomäne einen eigenen Simulator haben, der eine eigene Pipeline an Simulations-Stages verwaltet. Es kann theoretisch beliebig viele Simulationsdomänen geben. Momentan sind jedoch nur drei über einen `enum` definiert: Die visuelle, mechanische und akustische Domäne, wobei die akustische Domäne noch nicht in der prototypischen Implementierung vorkommt. Abb. 4 zeigt grob beispielhaft den Ablauf einer Simulation und die verschiedenen Varianten, wie man sich die „Rendering Results“ anderer Stages beschaffen kann:

Entweder aus der Main Loop der Engine heraus oder von der benutzenden Applikation wird `URE::stepSimulation()` aufgerufen; Diese Funktion iteriert in fester Reihenfolge über die vorhandenen Simulatoren und ruft deren `stepSimulation()`-Routine auf; Hierin werden Simulator- spezifische Operationen ausgeführt (Framebuffers des Fensters resetten und Haupt-Viewport setzen beim `LightingSimulator`, CL/GL-shared Buffers für OpenCL akquirieren beim `MechanicsSimulator...`), dann über die einzelnen zugehörigen `SimulationPipelineStage`'s iteriert, widerum deren `stepSimulation()`-Methode aufgerufen. Hier passiert dann das eigentlich *generische Rendering*. So verschieden die Algorithmen sind, die in den einzelnen Stages abgearbeitet werden, so verschieden sind auch die Möglichkeiten, wie einzelne Stages miteinander Daten austauschen: Der generischste Ansatz ist, sich von einer Stage einen Buffer mit einer bestimmten Semantik zu beschaffen, sofern die entsprechende Stage ein solchen bereit stellt. Bei Lichtsimulation, die viel mit Texturen und OpenGL-Framebuffer-Objects (FBOs) umgeht, bietet sich an, sich das `RenderTarget` (die Abtraktion des FBO in `Flewnit`) zu beschaffen, um mit dem gesamten RenderTarget weiter zu arbeiten, oder sich einzelne Texturen zu beschaffen. Es gibt aber auch die relativ indirekte Methode des Datenaustausches, wie es z.B. zwischen der `ParticleMechanicsStage` und der `ParticleLiquidDrawStage` der Fall ist: Die Buffer mit dem physikalischen Attributen sind in der `SceneNode`-abgeleiteten Klasse `ParticleFluid` enthalten,

³²<http://www.antisphere.com/Wiki/tools:anttweakbar>

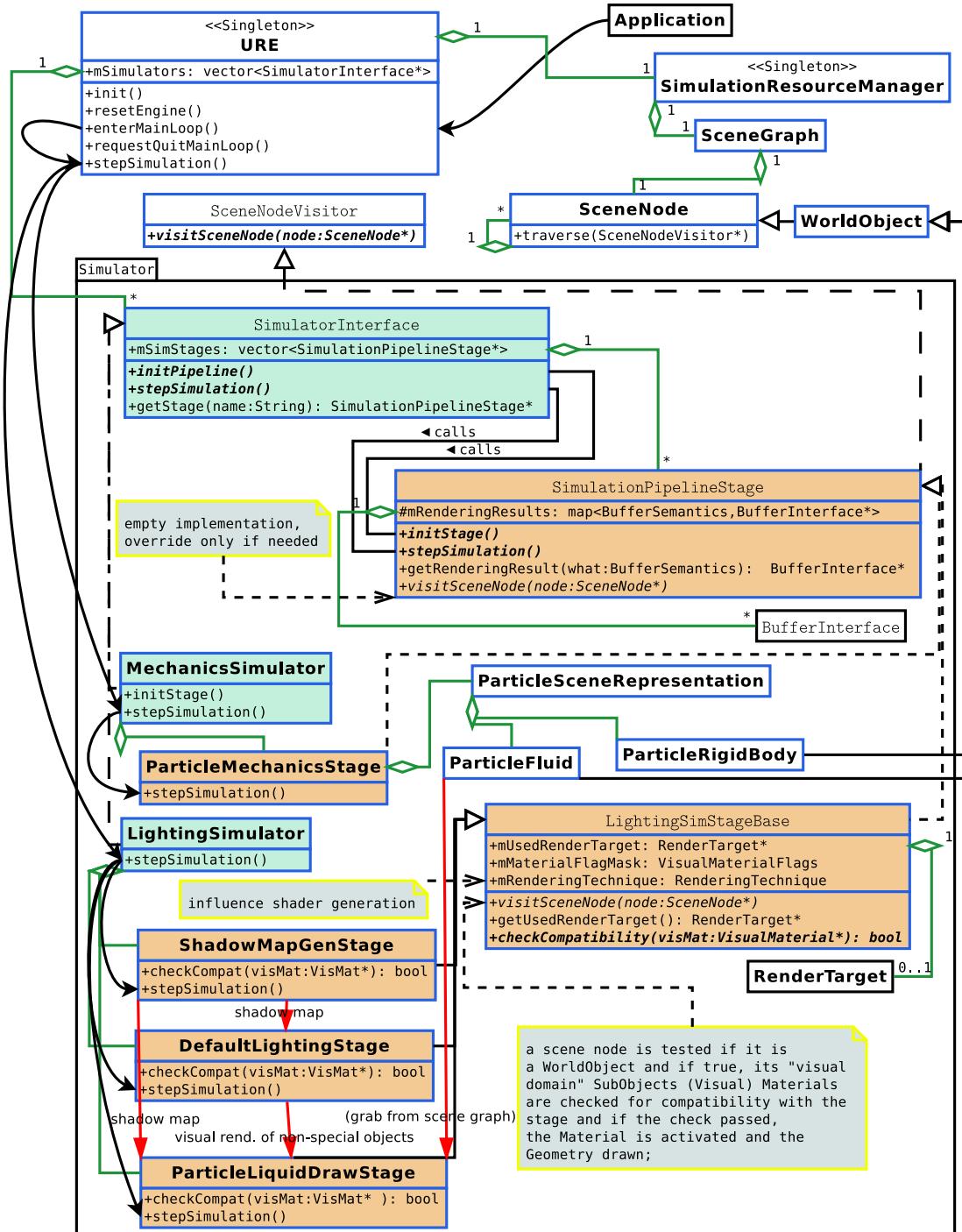


Abbildung 4: Grobes Beispiel von einer Simulation und den zugehörigen Daten-Abhängigkeiten

so dass die aktualisierten Daten über schlichtes Szenegraphen-Traversieren beschafft werden können.

Details zum Ablauf der Simulation(en) werden in Kapitel 4 behandelt.

3.8 Der SimulationResourceManger

Diese Singleton-Klasse besitzt und verwaltet die Assets, als `std::map<String,X*>` für Referenzierung über einen Namen, wobei x folgende Klassen betrifft:

- Material
- Geometry
- BufferInterface
- Texture³³
- InstanceManager

Außerdem bestitzt die Klasse den `SceneGraph` und stellt einen Handle auf den aktuellen `SkyDome` zur Verfügung, so dass alle Objekte ein konsistentes Environment Mapping betreiben können, sofern dies erwünscht ist.

3.9 Das Scene-Paket

Neben dem klassischen Szene-Graphen, der vor allem im Kontext des visuellen Echtzeit-Rendering verwendet wird, da er Culling und relative Transformationen ermöglicht, erfordern verschiedene Simulationsdomänen und Repräsentationen der beteiligten Simulations-Objekte (z.B. als Dreiecks-Mesh, Komposition von komplexeren Primitiven, Punktfolge oder Voxel-Grid) ggfs. unterschiedliche Szene-Repräsentationen, damit die spezifischen Anforderungen der Simulation als solchen und der Repräsentation der simulierten Objekte optimal erfüllt sind.

Diese Szenen-Repräsentationen unterscheiden sich in der logischen Organisation ihrer Objekte (z.B. Baumstruktur, sortierte oder unsortierte Sammlung, Graph-Struktur, Voxel-Struktur) und in den Meta-Informationen, welche für die einzelnen Objekte notwendig sind.

Beispielsweise stellt die die effiziente Suche nach räumlich benachbarten Objekten eine wichtige Anforderung bei vielen Simulationen (auch bei Lichtsimulation für globale Beleuchtungseffekte) dar. Hierfür bieten sich verschiedene Beschleunigungsstrukturen an. Je nach Simulationsdomäne, Objekt-Repräsentation, verwendetem Simulations-Verfahren, verwendeter Hardware

³³Da Texture von BufferInterface erbt, handelt es sich um eine Handle-Sammlung der Untermenge der Texturen aller BufferInterfaces, um einen bequemerem Zugriff zu ermöglichen

	Szenen-Repräsentation	Beschleunigungsstruktur
Organisation der Elemente	logisch für Zugriff auf Ebene der Anwendungslogik	räumlich für effizienten Zugriff durch einen Simulations-Algorithmus
Zweck von Meta-Informationen	verschieden, je nach Repräsentation und Typ der Simulations-Objekte	(sofern vorhanden) unterschiedlich für verschiedene Algorithmen, z.B. effiziente Traversierung und/oder Nachbarschaftssuche

Tabelle 1: Gegenüberstellung der Zwecke einer generischen Szenen-Repräsentation und einer Beschleunigungsstruktur

und Zielsetzung (Geschwindigkeit, Genauigkeit, Speicherverbrauch...) sind verschiedene Beschleunigungsstrukturen angemessen. Somit besteht keine 1:1-Verbindung zwischen Szenen-Repräsentation und Beschleunigungsstruktur.

Beschleunigungsstruktur und Szenen-Repräsentationen sind somit nur lose gekoppelt. Um die verschiedenen Zwecke zu verdeutlichen, stellt Tabelle 1 diese gegenüber.

Es gibt also zwei Basis-Klassen, wie in Abb. 3 zu sehen ist: `SceneRepresentation`↔ und `AccelerationStructure`. `AccelerationStructure` erbt von `WorldObject` (siehe Abschnitt 3.10), um optional Debug-Draw-Funktionalität bereit zu stellen. Ansonsten handelt es sich schlicht um abstrakte Basis-Klassen ohne weitere Funktionalität.

Für das gesamte System über den `SimulationResourceManger` verfügbar ist der `SceneGraph`. Dieser enthält die Wurzel-`SceneNode`, welche beliebig viele Kinder derselben Klasse haben kann. Somit wird eine Baumstruktur gebildet. Die `SceneNode` bildet die Basis-Klasse des `WorldObject` (siehe Abschnitt 3.10).

Es ist ein Mechanismus konzipiert, sowohl Kinder als auch Vorfahr über Änderungen der eigenen Transformation und/oder Bounding Box zu informieren, so dass ggfs. globale Transformationen akkumuliert werden können und/oder die Bounding Box für den eigenen Unterbaum ermittelt, die für Culling-Berechnungen verwendet werden kann. Noch ist dieses Feature nicht in Nutzung und seine Implementation unvollständig und ungetestet, daher wird auf weitere Details verzichtet.

Die klassischen Operationen (Einfügen, Aushängen, rekursives Suchen im Unterbaum nach Namen, Node bewegen, Transformationen akkumulieren) sind jedoch implementiert. Es sei an die enge Kopplung dieser Klasse an `AmendedTransform` (s. Abschnitt ??) erinnert.

Der Scenegraph kann nach dem Visitor- Design Pattern traversiert wer-

den. Z.B. erbt `SimulationPipelineStage` von `SceneNodeVisitor`, so dass bei Bedarf eine Pipeline-Stage nur `virtual void visitSceneNode(SceneNode* node)` implementieren muss, um Operationen auf oder mit den Objekten des Szene-graphen auszuführen. Momentan „besitzt“ die `SceneNode`-Klasse seine Kinder, was den Vorteil hat, dass die Nodes nicht woanders verwaltet werden müssen, aber den Nachteil, dass ausgehängte Scene Nodes nicht automatisch beim Reset der Engine gelöscht werden. Sollte sich dies später als Problem herausstellen, könnte man die Scene Nodes auch vom `SimulationResourceManger` verwalten lassen. Auch empfiehlt sich langfristig die Möglichkeit zur Er-schaffung und Verwaltung mehrerer Szenenraphen. Aus Zeitgründen habe ich es jedoch erstmal bei dieser rudimentären Implementation belassen.

Intern verwaltet die `ParticleMechanicsStage`, die `SimulationPipelineStage`, die im `MechanicsSimulator` für die partikelbasierte Simulation zuständig ist, eine `ParticleSceneRepresentation`. Diese verwaltet

- die „makroskopischen Objekte“, die an dieser Simulation teilnehmen, namentlich (allesamt von `WorldObject` abgeleitete) partikelbasierte Flu-ide, partikel-repräsentierte Rigid Bodies und statische Kollisions-Meshes³⁴
- die einzelnen Attribute-Buffers der Partikel (Position, Dichte, Geschwindigkeit, Beschleunigung usw.)
- Buffers mit Meta-Informationen der „makroskopischen Objekte“, die an OpenCL-Kernels übergeben werden

Ferner stellt die Klasse Factory-Methoden für die makroskopischen Objekte bereit, da diese direkt mit den Partikel-Buffern assoziiert sind, sich also alle beteiligten Objekte die Buffers teilen müssen. Die korrekte Delegierung der Offsets und Sicherstellung der richtigen Größen wird so übernommen. Die `ParticleMechanicsStage` nutzt außerdem ein `UniformGrid` für die Simula-tion. Details hierzu werden im Verlauf des Kapitels 4.2 beschreiben.

3.10 Das WorldObject

Das `WorldObject` stellt die abstrakte Basisklasse für sämtliche Objekte dar, mit denen eine oder mehrere Simulationen durchgeführt werden sollen³⁵. Sie erbt von `SceneNode`. Damit ist sie bequem in den Szenenraphen integrier-bar und hat außerdem schon seinen Transformations-Membervariable.

Bei dieser Klasse zeigt sich erstmals die angestrebte Symmetrie zwis-chen sen Simulationsdömanen: Für jede Simulationsdomäne gibt es eine

³⁴diese müssen in einem bestimmten Format vorliegen

³⁵Und sei es nur Debug Drawing; Selbst dies kann als eine rudimentäre Form von Licht-simulation aufgefasst werden.

(möglicherweise leere) Liste an `SubObject`'s. Das `SubObject` stellt eine Abstraktion dessen dar, was in einer klassischen Graphik-Engine ein „Sub-Mesh“ oder in einer klassischen Physik-Engine eine Komponente einer „(Compound) Collision Shape“ sein könnte.

Ein `SubObject` hat immer einen Pointer auf genau ein `Material`, eine `Geometry` und für Backtracking-Zwecke auf sein besitzendes `WorldObject`. Wie der Backtracking-Pointer andeutet, besitzt das ein `WorldObject` jedes seiner `SubObject`'s. Die vom `SubObject` genutzte Geometrie und Materialien können jedoch –sofern sinnvoll – von beliebig vielen `SubObjects` gemeinsam genutzt werden.

Es sind verschiedenste Ableitungen von `WorldObject` denkbar, z.B. (standard/particle based/voxel based) Rigid Body, Soft Body, rein visuelles Objekt, Kamera, Lichtquelle, Debug-Draw-Objekte, Cloth, Hair, oder (particle based/voxel based) Fluid.

Welche Objekte im Szenengraph für die Nutzung in einer Simulations-Stage in Frage kommen, lässt sich entweder über eine Vorfilterung bei Objekt-Erstellung (Factory-Pattern oder automatische Registrierung bei einer Manager-Klasse im Konstruktor), oder über `dynamic_cast<>()`'s und anschließendes Auslesen von spezifischer Meta-Information ermitteln, oder aber durch „property flags“, also Bit-Flags, die Features einer Scene Node andeuten. Welche dieser Methoden langfristig die robusteste, flexibelste und eleganteste ist, vermag ich noch nicht zu sagen. Da Typ-Informationen über `enums` (als Bitflags oder „normale“ Aufzählungstypen) eine gewisse Redundanz hat ggü. der dynamischen Typ-Information, die C++ bereit stellt, und außerdem eine weitere Fehlerquelle für den Programmierer darstellt, sind die Bitflags nicht erste Wahl. Außerdem schränkt die explizite Auflistung von Features womöglich die Erweiterbarkeit des Systems ohne komplette Neu-Kompilierung ein. Andererseits machen sie Programme vielleicht lesbarer, indem bestimmte Kategorien/Features explizit aufgelistet (und im Falle von Bitflags beliebig kombinierbar) sind, und nicht implizit in C++-Typinformationen codiert sind. Ich verwende Aufzählungstypen intensiv in *Flewnit*. Nicht immer sind diese für die Programmlogik nötig. Sie halfen mir jedoch zur Strukturierung und Ideen-Sammlung. Letztere ist wichtig, da bei einem Softwaresystem, welches generische Verwendbarkeit anstrebt, schon weit vor Implementierung konkreter Klassen die Struktur der Basisklassen anhand von abstrahierten (erwarteten) gemeinsamen Features der möglichen abgeleiteten Klassen möglichst exakt bestimmt werden sollte. Je weiter oben in der Klassenhierarchie Design-Fehler auftauchen, desto aufwändiger droht das spätere Refactoring zu werden.

3.11 Material

Der Begriff „Material“, der sich in der Computergraphik eingebürgert hat als Beschreibung von visuellen Eigenschaften einer Oberfläche, erscheint

mir prädestiniert, sich in seiner Bedeutung in einem alltäglichen, nicht computergraphischen Kontext wieder anzunähern: das *Material* in *Flewnit* steht für „sämtliche Eigenschaften von Materie außer ihrer geometrischen Form und internen Repräsentation dieser“.



Abbildung 5: Der Begriff *Material* mit einer physikalisch orientierten Bedeutung: sämtliche Eigenschaften von Materie außer ihrer geometrischen Form. (Bild entnommen aus [APS00])

Würden wir über die entsprechende Rechenleistung verfügen und als Programmierer das physikalische Know-How besitzen (z.b. über Maxwell'sche Gleichungen), bräuchte man dieses Konzept gar nicht weiter zu konkretisieren, und könnte sämtliches Verhalten von Simulations-Objekten durch subatomare Material-Eigenschaften modellieren. Dies wäre eine "wirkliche" Vereinheitlichung der bestehenden Engine-Konzepte, die unserer Realität Rechnung trüge. Da wir jedoch realistisch (im Sinne der Machbarkeit) bleiben wollen jetzt und heute Echtzeit-Fähigkeit, auch auf Kosten von physikalischer Korrektheit/Genauigkeit anstreben, sind Domänen-spezifische Konkretisierungen des Material-Konzeptes notwendig, und wie so viele andere Klassen in diesem System wird die `Material`-Klasse zur abstrakten Oberklasse. Bei einem `SubObject`, welches der visuellen Simulationsdomäne angehört, wird implizit erwartet, dass das genutzte `Material` ein `VisualMaterial` ist (mit Texturen, Shader, und verschiedenen Meta-Informationen, um das visuelle Rendering zu delegieren, siehe Abschnitt 4.1.3), und in der mechanischen Domäne muss es ein `MechanicalMaterial` (mit Eigenschaften wie Masse, Reibung, Elastizität etc.) sein. Diese domänen-spezifischen Material-Typen können oder müssen (je nach Typ des nutzenden `WorldObjects`) noch weiter abgeleitet sein.

Einen etwas schalen Beigeschmack haben die Routinen:

```

1 virtual void activate(
2     SimulationPipelineStage* currentStage,
3     SubObject* currentUsingSubobject) throw(SimulatorException){}
4 //undoing stuff, like re-enable depth test etc.
5 virtual void deactivate(SimulationPipelineStage* currentStage,
6     SubObject* currentUsingSubobject) throw(SimulatorException){}

```

Sie sind dem Statemachine-basierten visuellen Rendering mit OpenGL zu verdanken und müssen daher von den `VisualMaterial`-Klassen implementiert werden. Um der Möglichkeit Rechnung zu tragen, dass ähnlich State-basierte Mechanismen an anderer Stelle als beim OpenGL-Rendering vorkommen sollten, sind diese Routinen in die Oberklasse gelangt.

3.12 Die Buffer-Abstraktion

Logischerweise müsste nun eigentlich – der Top-Down-Präsentation der einzelnen Pakete und Klassen entsprechend – die `Geometry`-Klasse vorgestellt werden. Um letztere Klasse jedoch besser zu verstehen, ist die Kenntnis der Buffer-Abstraktion vorteilhaft. Die `Geometry` wird in 3.13 beschrieben.

Wie auf Seite 12 erwähnt, abstrahiert die `BufferInterface`-Klasse sämtliche Buffer-bezogene Funktionalität, namentlich für Host-Buffer, OpenCL-Buffer, verschiedene OpenGL-Buffertypen und letztendlich verschiedene Texturtypen (eine Teilmenge davon ist OpenCL-interop-kompatibel bzw. als reine OpenCL-Textur erschaffbar). Es ergibt sich ein wahrer Moloch, der für eine relativ einheitliche Handhabung zu abstrahieren ist:

- verschiedene APIs (OpenGL C-API „vs.“ OpenCL C++-API)
- verschiedene Verwendungen und Verfügbarkeiten schon allein von non-Textur-Buffern (generisch, Vertex Index/Vertex Attribute Buffer, Uniform Buffer, Render Buffer), siehe Tabelle 2
- verschiedene OpenGL-Texturtypen mit nur bedingt kombinierbaren und nur teilweise zu OpenCL kompatiblen erweiterten Features, siehe Tabelle 3
- verschiedene Datenformat/Channel-Layout-Deskriptoren, siehe Listing 7

Es hat schon sehr viel Zeit gekostet, sich aus den Spezifikations-Dokumenten ([Khr10], [Khr09]) die möglichen Permutationen, Zusammenhänge, Entsprechungen und unterstützten Operationen zu erarbeiten (siehe Tabellen 2 und 3).

Die Ergebnisse dieser Tabellen flossen in die Entscheidung ein, welche konkreten Buffer- und Textur-Klassen mit welchen verfügbaren Features durch Ableitung vom `BufferInterface` implementiert wurden. Die Konstruktor-Parameter sollten so spezifisch zugeschnitten sein, dass der Benutzer der Klassen möglichst wenig invalide Kombinationen angeben kann. Außerdem sollte er sich nicht um all die Makros wie `GL_RGBA32UI` bzw `CL_RGBA` und `CL_UNSIGNED_INT32` kümmern müssen³⁶. Es soll ein Buffer mit einem einzi-

³⁶Was in OpenCL getrennt ist und somit „nur“ 4 Channeltypen + 3 Bitgrößen * (3 non-normalized + 2 normalized)= theoretisch 19 makros ergibt, braucht bei OpenGL 4 Channeltypen * 3 Bitgrößen * (3 non-normalized + 2 normalized)= theoretisch 60 makros; Erstens stört die Asymmetrie, zweitens das „Zusammenklatschen“ eigentlich unabhängiger Parameter in ein Makro

gen Konstruktor-Call vollständig definiert, falls erwünscht auf dem Host-Memory und – sofern unterstützt – in den gewünschten API-Contexten allokiert bzw registriert werden.

Hierfür mussten einige `enums` und Meta-Info-Klassen/Structures definiert werden:

enum Type Beschreibt String-, boolsche, Skalar-, Vektor- und Matrix-Datentypen mit verschiedener Genauigkeit. Findet in *Flewnit* u.a. auch beim Config-Parsen Verwendung.

enum ContextType und ContextTypeFlags Mithilfe dieser Datentypen lässt sich spezifizieren, in welchem Kontexten (Host, OpenGL, OpenCL) man einen Buffer erstellt/allokiert/registriert haben will.

enum BufferSemantics Eine sehr wichtige Auflistung konkreter Verwendungszwecke des Buffers, siehe Listing 5 im Anhang A. Dieser Aufzählungstyp ist eine große Erleichterung beim Hantieren mit OpenGL Vertex Buffer Objects (VBO's)³⁷ und Frame Buffer Objects (FBO's)³⁸, da hier zur Assoziation von VBO-Vertex Attribute Buffern mit Input-Variablen im Vertex Shader bzw ans FBO gehängte Texturen mit Output-Variablen im Fragment Shader ein Index angegeben werden muss. Wenn wir als Index die numerische Entsprechung der Buffer-Semantik verwenden, ist nicht nur die Lesbarkeit des Codes erhöht, es ist auch automatisch sichergestellt, dass Indizes keine Konflikte versursachen!

enum GPU_DataType Listet nur die von der GPU / OpenGL nativ unterstützten Datentypen auf, ohne Präzisions- und Normalisierungs-Information:

```
1 enum GPU_DataType
2 {
3     GPU_DATA_TYPE_FLOAT,
4     GPU_DATA_TYPE_INT,
5     GPU_DATA_TYPE_UINT
6 }
```

enum GLBufferType Listet die vom Buffer-Interface abstrahierten OpenGL non-Textur-Buffertypen auf. Aus Zeitgründen und Erwartung, dass sie außerhalb der `RenderTarget`-Klasse nicht benötigt werden, werden z.B. Render Buffers nicht abstrahiert.

```
1 enum GLBufferType
2 {
```

³⁷abstrahiert von `Flewnit::VertexBasedGeometry`, s. 3.13

³⁸abstrahiert von `Flewnit::RenderTarget`

```

3     NO_GL_BUFFER_TYPE,
4     VERTEX_ATTRIBUTE_BUFFER_TYPE,
5     VERTEX_INDEX_BUFFER_TYPE,
6     UNIFORM_BUFFER_TYPE
7 };

```

struct BufferElementInfo Beschreibt – sofern erwünscht bzw nötig – das Channel-Layout, die Datentypen, die Bit-Genauigkeit und die etwaige Normalisierung eines Buffers mit Channels, z.B. einer Textur oder einem Vertex Attribute Buffer, siehe Listing 7 im Anhang A. Auf diese Weise spart man dem Benutzer die verschiedenen „zusammengesklatschten“, asymmetrischen GL/CL-Makros; Die Structure validiert ihre Daten. Diese werden später für die internen CL/GL-API-Aufrufe in die entsprechenden CL/GL-Makros transformiert, siehe Listing 8 im Anhang A.

struct GLImageFormat OpenCL definiert folgende Structure:

```

1 typedef struct _cl_image_format {
2     cl_channel_order      image_channel_order;
3     cl_channel_type       image_channel_data_type;
4 } cl_image_format;

```

Um das interne Handling der CL/GL-Makros etwas symmetrischer zu machen (man verliert leicht den Überblick), habe ich ein ähnliches Makro definiert, welches aufgrund der Kommentare im Anhang A, Listing 6 zu finden ist.³⁹

class BufferInfo Dies ist die Klasse, die sowohl als „Construction Info“ den Konstruktor-Code delegiert, als auch für den Benutzer später Meta-Information bereitstellt. Es sei hier nur der Standard-Konstruktor angegeben:

```

1 explicit BufferInfo(String name,
2   ContextTypeFlags usageContexts,
3   BufferSemantics bufferSemantics,
4   Type elementType,
5   cl_Gluint numElements,
6   const BufferElementInfo& elementInfo,
7   GLBufferType glBufferType = NO_GL_BUFFER_TYPE,
8   ContextType mappedToCPUContext = NO_CONTEXT_TYPE);

```

Ich muss zugeben, dass `elementType` und `elementInfo` redundant sind, außerdem diese beiden Parameter eigentlich bei generischen Buffern,

³⁹Die Kommentare geben weiteren Aufschluss über die Asymmetrie der einzelnen APIs: obwohl sie beide von der Khronos Group spezifiziert wurden und explizit Interoperabilität ermöglichen, stifteten viele mehr oder minder kleine Asymmetrien Verwirrung (zumindest bei mir); Diese Umstände trugen zu der Entscheidung bei, das `BufferInterface` zu entwickeln, trotz des erheblichen Zeitaufwands.

wo die Typ-Information zumindest für die GPU Computing APIs keine Rolle spielt, fehl am Platze sind. Mangelndem Überblick in der Anfangsphase der Implementierung über die verschiedenen Arten, Buffers in CL und GL zu nutzen, sind diese Design Flaws zuzuschreiben. Ein Refactoring mit spezieller zugeschnittenen Konstruktoren ist geplant.

class TextureInfo So wie `Texture` von `BufferInterface` erbt, erbt auch `TextureInfo` von `BufferInfo`. Sie stellt die gleiche Funktionalität für die (abstrakte) `Texture`-Klasse dar wie `BufferInfo` für `BufferInterface`. Ein Unterschied besteht jedoch darin, dass `TextureInfo` eigentlich nur dazu dient, die kombinierte Information aus den Parametern der „eingeschränkten“ Konstruktoren der konkreten Textur-Klassen (`Texture2D` etc.) und der inhärent klassenspezifischen Eigenschaften an die `Texture`-Oberklasse zu übergeben. Damit hat `TextureInfo` eher eine Textur-interne Funktion bei der Konstruktion. Das tut jedoch seiner Funktion als späterer Meta-Daten-Lieferant keinen Abbruch. Auch hier reicht der Standard-Konstruktor für eine Idee dieser Klasse:

```

1 explicit TextureInfo(
2     const BufferInfo& buffi,
3     cl_GLuint dimensionality,
4     Vector3Dui dimensionExtends,
5     GLenum textureTarget,
6     bool isDepthTexture = false,
7     bool isMipMapped = false,
8     bool isRectangleTex = false,
9     bool isCubeTex = false,
10    GLint numMultiSamples = 1,
11    GLint numArrayLayers = 1
12 );

```

Es sei bemerkt, dass die Tiefen-Texturen sich abgesehen von den automatisch bestimmten entsprechenden OpenGL-Makros in ihrer Implementation nur geringfügig von ihren „Farbtextur“-Oberklassen unterscheiden, nämlich in der Implementation der `virtual void setupDefaultSamplerParameters(=0)`-Methode, die in `Texture` definiert ist. Hier wird z.B. der Compare-Mode und die Compare-Function für korrekten Lookup solcher Texturen als Shadow Map über einen Shadow-Sampler in GLSL gesetzt.

Man darf sich fragen, warum ich jede noch so sonderbar anmutende Spezial-Textur unterstützen will. Dies hat den Grund, dass ich langfristig etliche Rendering Features unterstützen will, welche die meisten dieser Spezialtexturen benötigen, und die verbleibenden paar ungenutzten Texturen noch zur Vollständigkeit mit zu implementieren, verursacht dank Vererbung nur einen minimalen Mehraufwand:

1. Layered Rendering in `Texture2DDepthCube` zur Generierung von Point Light Shadow Maps in nur einem Rendering Pass

		Context		
		Host	OpenGL	OpenCL
generic Buffer		✓	✗	✓
OpenGL Buffers	Vertex Attribute Buffer	○	✓	○
	Vertex Index Buffer	○	✓	○
	Uniform Buffer	○	✓	○
	Render Buffer	✗	✓	✓
Textures	1D Texture	○	✓	✗
	2D Texture	○	✓	✓
	3D Texture	○	✓	✓
	Special Texture	?	✓	?

Tabelle 2: Verschiedene Buffertypen und ihre Verfügbarkeit in verschiedenen Kontexten

Legende:

✓ → nativ unterstützt; ○ → kompatibel; ✗ → nicht unterstützt;
? → Unterstützung abhängig von weiteren Parametern, s. Tabelle 3;

	CL interop	MipMap	Depth	Array	Rectangle
Texture1D	✗	✓	✗	✓	✗
Texture2D	✓	✓	✓	✓	✓
Texture2DCube	○	✓	✓	○	✗
Texture2DMultiSample	✗	✗	✗	✓	✗
Texture3D	✓	✓	✗	✗	✗

Tabelle 3: Verschiedene Texturtypen und ihre Kompatibilität zu bestimmten Features

Legende:

✓ → unterstützt; ○ → In OpenCL unterstützt, aber nicht vom Framework; ○ → nur in OpenGL 4 unterstützt, daher wegen Kompat. zu GL 3 nicht vom Framework unterstützt; ✗ → nicht unterstützt;

2. Layered Rendering in `Texture2DCube` zur Generierung von dynamischen Environment Maps
3. Layered Rendering in `Texture2DDepthArray` zur Generierung von vielen Spot Light Shadow Maps in nur einem Rendering Pass
4. Deferred Shading mit Multisampling: Benötigt `Texture2DMultiSample` -G-Buffers

Zu weiten Teilen sind der Shadercode und die nötigen Framework-Klassen schon für diese Features implementiert, es muss „nur“ noch einiges ergänzt und aufgeräumt, alles zusammengefügt und debuggt werden. Da dieses „Nur“ wirklich äußerst ironisch gemeint ist, habe ich gar nicht erst versucht, die Implementierung dieser Features zu beenden. Es muss vorerst

reichen, dass sie konzeptionell in der Framework-Struktur angelegt sind.

Alle Operationen auf den Buffern/Texturen durch den Benutzer werden dann über das `BufferInterface` getätigkt, siehe das exemplarische Listing 2. Fast sämtlicher Kontrollfluss (Synchronisation, Validierung, Auswahl des richtigen Kontextes etc.) befindet sich in der Implementation der (wohlge-merkt großteils *nicht* virtuellen) Methoden von `BufferInterface`. Nur die finale Bürde der verschiedenen API-Calls als Ergebnis des vorangegangenen Kontrollflusses wird von `protected` virtuellen Funktionen der abgeleiteten Klassen übernommen, wie im folgenden Unterabschnitt beschrieben.

Listing 2: Operationen auf dem BufferInterface, Ausschnitt

```
1 //The binding-mechanism is only relevant for OpenGL
2 void bind() throw(BufferException);
3 void setData(const void* data, ContextTypeFlags where) throw(←
4     BufferException);
5 //if both CL and GL are enabled, then the buffer is shared and the ←
6 //implementation
7 //will decide which API will be used for the read/write;
8 void copyFromHostToGPU() throw(BufferException);
9 void readBack() throw(BufferException);
```

3.12.1 Die API-bezogenen internen Buffer-Operationen

Zur Vermeidung von Boilerplate-Code sind die Signaturen für die GPU Computing API-bezogenen Buffer-Operationen ausgelagert in eine Datei namens „`BufferVirtualSignatures.h`“. Um sowohl in der abstrakten Basisklasse als auch in den abgeleiteten Klassen included werden zu können, lässt sich die Abstraktheit der Definitionen über das Makro `FLEWNIT_PURE_VIRTUAL` steuern:

Listing 3: API- und Buffertyp-abhängige Operationen auf Buffern – Definitionen

```
1 #ifdef FLEWNIT_PURE_VIRTUAL
2 # define PURENESS_TAG =0
3 #else
4 # define PURENESS_TAG
5 #endif
6
7     virtual void generateGL() PURENESS_TAG;
8     virtual void generateCL() PURENESS_TAG;
9     virtual void generateCLGL() PURENESS_TAG;
10
11    //there is no pendant to bind() in OpenCL, as the API needs no buffer ←
12    //binding mechanism;
13    //instead, buffers are passed as arguments to kernels
14    virtual void bindGL() PURENESS_TAG;
15    //there is no pendant to alloc() in OpenCL, as Buffer generation and ←
16    //allocation are coupled within
17    //the same API call
```

```

16 virtual void allocGL() PURENESS_TAG;
17
18 virtual void writeGL(const void* data) PURENESS_TAG;
19 virtual void writeCL(const void* data) PURENESS_TAG;
20 virtual void readGL(void* data) PURENESS_TAG;
21 virtual void readCL(void* data) PURENESS_TAG;
22 virtual void copyGLFrom(GraphicsBufferHandle bufferToCopyContentsFrom) ←
    PURENESS_TAG;
23 virtual void copyCLFrom(ComputeBufferHandle & bufferToCopyContentsFrom) ←
    PURENESS_TAG;
24 //because the C++ Wrapper for OpenCL is used, deletion takes place ←
    automatically when the cl::Buffer is destroyed
25 virtual void freeGL() PURENESS_TAG;
26
27 // //not implemented yet as not needed at the moment: mapping routines:
28 // virtual void* mapGLToHost() PURENESS_TAG;
29 // virtual void* mapCLToHost() PURENESS_TAG;
30 // virtual void* unmapGL() PURENESS_TAG;
31 // virtual void* unmapCL() PURENESS_TAG;
32
33 #undef PURENESS_TAG

```

So included `BufferInterface` diese Datei innerhalb ihrer Klassendefinition über

```

1 # define FLEWNIT_PURE_VIRTUAL
2 # include "BufferVirtualSignatures.h"
3 # undef FLEWNIT_PURE_VIRTUAL

```

Die anderen Klassen definieren das Makro nicht vor dem Includen.

In Anhang A sind exemplarisch Implementationen dieser Funktionen gelistet, einerseits die der konkreten `Buffer`-Klasse (Listing 9), die alle non-Textur-Buffer zusammengefasst abstrahiert, andererseits die der konkreten `Texture2D`-Klasse (Listing 10 und 11), die von der abstrakten `Texture`-Klasse erbt und selbst die Basisklasse für `Texture2DDepth` und `Texture1DArray` bildet, vgl. Klassendiagramm (Abb. 3). Die meisten Signaturen aus „`BufferVirtualSignatures.h`“ werden schon von `Texture` implementiert, da nicht jede API-Routine Textur-spezifisch ist.

Ich muss gestehen, dass ich diese Funktionen noch nicht systematisch getestet habe, daher Fehlerfreiheit nicht garantiert ist. Es sei nochmals betont, dass der Schwerpunkt bei einer möglichst generischen Konzeption lag, nicht bei einer vollständigen Implementation. Es ist viel „schlafender“ Code im System, der noch getestet, integriert, weiter entwickelt, ergänzt werden will.

3.12.2 PingPongBuffer

Auch die für viele Algorithmen aufgrund ihrer inhärenten Struktur oder aufgrund der Beschränkungen bei der Synchronisation parallelen Codes

nötige „Ping-Pong“-Funktionalität⁴⁰ ist durch das `BufferInterface` abstrahiert: Der `PingPongbuffer` verwaltet intern zwei andere `BufferInterface`-Objekte und leitet die Methoden-Aufrufe, die durch das `BufferInterface` definiert sind, an den aktuell aktiven Buffer weiter. Der Rollentausch geschieht dann über `PingPongBuffer::toggleBuffers()`. Bei komplexeren Algorithmen, bei denen viele Toggles passieren, kann man schnell den Überblick verlieren, welcher Buffer jetzt gerade aktiv ist; Besonders kritisch ist dieser Umstand dort, wo ein OpenGL Buffer-Handle gebunden ist, oder ein OpenCL-Buffer-Handle als Argument für einen OpenCL-Kernel gesetzt ist. Diese Bindings lassen sich nicht von einem `toggleBuffers()` beeindrucken, da die `BufferInterface`-Klasse selbst nicht weiß, in welche Hierarchie an verschiedenen Binding-Points seine verwalteten Handles gerade integriert sind oder nicht. Um dem Programmierer die Bürde zu nehmen, neben den Toggles auch noch ständig die CL/GL-Bindings zu ändern, wurden Mechanismen implementiert, die diese Operationen automatisch tätigen, nämlich bisher in `VertexBasedGeometry`↔ um Attribute und Index Buffer Bindings konsistent zu halten (s. Abschnitt 3.13.2) und in `CLKernelArgument` (s. Abschnitt 3.14.3.1), um die Kernel-Arguments nicht „von Hand“ immer neu setzen zu müssen.

3.13 Geometry

Die `Geometry`-Klasse bildet die abstrakte Basisklasse verschiedenster geometrischer Reräsentationen.

Beispiele solcher Repäsentationen sind:

- Vertex-basiert: Punkte, Linien(-Züge), Dreiecke, Quads und beliebige Polygone lassen sich theoretisch aus ihnen konstruieren⁴¹
- Voxel-basiert
- Freiform-Flächen-basiert: NURBS, Bézier-Patches etc.
- Primitiv-Basiert: das Objekt ist aufgebaut aus Primitiven die Quader, Kugel, Kapseln, Zylindern etc.; Diese Repräsentation wird gerne für Collision Shapes bei der mechanischen Simulation von Rigid Bodies verwendet.

Die Repräsentationen Überschneiden sich in ihren Kategorisierungen und Eigenschaften, eine sinnvolle Strukturierung ist also weder trivial noch womöglich eindeutig. Die API-Spezifischen Beschränkungen erleichtern diese Aufgabe nicht gerade, wenn man das Effiziente Mapping auf eine API im Hinterkopf behalten muss. Ich habe mir nicht weiter den Kopf zerbrochen,

⁴⁰also die Verwendung von zwei Buffern, einem zum Lesen, einem zum Schreiben, und die Rollen werden nach jedem Schritt getauscht

⁴¹theoretisch deshalb, weil z.B. OpenGL maximal Dreiecks-Primitive unterstützt, im Falle von Tessellation noch Quads, aber das ist buchstäblich ein anderes Kapitel

auch, weil mir spezifisches Wissen z.B. über Primitiv-basierte Rigid Body-Simulation fehlt, und habe mich pragmatisch in der exemplarischen Implementierung auf die beiden Basis-Repräsentationen beschränkt, die bei aktuellem visuellen Echzeitrendering und interaktiver Fluidsimulation die wichtigste Rolle spielen: Vertex- und Voxel-basiert. Hierbei sind auch API-spezifische Überlegungen mit eingeflossen. Man muss abwägen zwischen der konzeptionell saubersten und umfassendsten Lösung und der effizienten Realisierbarkeit mit verfügbaren Hard- und Software-Architekturen; im Zweifelsfalle siegt das Konzept, was die effiziente Realisierbarkeit begünstigt.

Die Basisklasse hat die Methode

```
1 virtual void draw(
2     unsigned int numInstances=1,
3     GeometryRepresentation desiredGeomRep = DEFAULT_GEOMETRY_REPRESENTATION
        ) = 0;
```

`numInstances` findet beim Hardware-Instancing Verwendung (s. Abschnitt 4.1.8.3), `desiredGeomRep` bietet die Möglichkeit – sofern kompatibel – für den aktuellen Draw-Call eine andere Repräsentation zu wählen. So lassen sich z.B. gezielt bestimmte Dreiecks-Meshes als Linen oder Punkte zeichnen, ohne dass andere Objekte beeinflusst würden, wie es bei einer globalen Einstellung der Fall wäre (z.B. über `glPolygonMode(GL_FRONT_AND_BACK, GL_LINE)`).

3.13.1 BufferBasedGeometry

Diese Klasse ist ebenfalls noch abstrakt; Sie ist die Basis für alle Klassen, deren geometrische Repräsentation ein 1:1-Verhältnis zwischen all ihren Attributen aufweist, also keine komplexen Datenstrukturen benötigt. Abgeleitet hiervon sind `VertexBasedGeometry` (siehe Abschnitt 3.13.2) und `VoxelGridGeometry` (nicht implementiert).

`BufferBasedGeometry` hat ein Array von `BufferInterface`-Pointern als Member, welches über den `BufferSemantics`-Aufzählungstyp indiziert werden kann:

```
1 BufferInterface* mAttributeBuffers[←
    __NUM_VALID_GEOMETRY_ATTRIBUTE_SEMANTICS__];
```

Diese Buffers können über

```
1 virtual void setAttributeBuffer(BufferInterface* buffi);
```

gesetzt werden. Die Position in `mAttributeBuffers` wird von der Semantik von `buffi` bestimmt. Die Methode ist virtuell, da sie von `VertexBasedGeometry`

für das Setzen des entsprechenden OpenGL-States überschrieben werden muss.

Zur Information, falls man mit dem System eine Voxel-basierte Simulation plant:

`VoxelGridGeometry` könnte als konkrete Attribute-Buffers entweder 3D-Texturen oder einen generischen Buffer haben. Hier tut sich ein Problem auf: Das Schreiben in 3D-Texturen ist zwar als OpenCL-Extension verfügbar, wird aber zumindest noch nicht von Nvidia-Treibern unterstützt, evtl. noch nicht einmal nativ von der Hardware (hierüber habe ich keine Informationen gefunden). Folglich würde man einen generischen Buffer verwenden wollen. Dieser lässt sich jedoch nicht mehr direkt per OpenGL via Ray-Casting visualisieren. Abhilfe könnte schaffen, dass man eine Textur allokiert, diese jedoch an OpenCL als klassischen Buffer übergibt. Derartige Mechanismen stellt OpenCL jedoch nicht bereit. In CUDA ist es zumindest möglich, einen generischen Buffer als Textur zu binden, darüber, ob es umgekehrt auch geht, habe ich keine Information⁴². 3D-Image Writes sind auch in CUDA noch nicht möglich. Man scheint sich also immer noch mit Work-Arounds behelfen zu müssen, wie 3D-Texturen in 2D-Texturen zu encoden oder pro Frame von einem Buffer in eine 3D-Textur zu kopieren.

Auch wenn das beschriebene Phänomen meine partikelbasierte Fluidsimulation nicht betrifft, finde ich es doch bemerkenswert, dass bei all den Fortschritten, die Hardware und APIs gemacht haben, teils immer noch Beschränkungen wie diese existieren.

3.13.2 VertexBasedGeometry

Dies Klasse abstrahiert das OpenGL Vertex Buffer Object(VBO).

Wie in den Textur- und Buffer-Klassen selbst, wird auch hier die `BufferInfo`-Klasse genutzt, um die relevanten Parameter (Datentyp und dessen Bit-Count, Anzahl Channel-Komponenten, Normalisierungs-Flags, BufferSemantics als Attribute Index, s. Seite 34) der VBO-bezogenen API-Calls zu bestimmen .

Die Klasse hat zusätzlich zu den Attribute Buffers einen Pointer, der optional als Vertex Index Buffer gesetzt werden kann. Dann wird der OpenGL-Draw Call über `glDrawElementsInstanced(..)` getätig. Zusätzlich kann man optional ein Intervall angeben, welcher Bereich im Index Buffer genutzt werden soll. Dies ist z.B. sinnvoll, wenn mehrere logische Objekte mit verschiedenen Materials in einem Buffer liegen, wie es z.B. bei einer Fluidsimulation der Fall ist, in welcher mehrere Fluide und Rigid Bodies sich ein Attribute Buffer Set teilen. So kann man kontrollieren, was im Buffer gezeich-

⁴²Texturen liegen intern so im Speicher, dass eine gute 2D-räumliche Lokalität gewährleistet ist, z.B. über eine Z-Order-Curve (siehe [?]). Auf diese Weise werden Cache Misses und benötigte Bandbreite reduziert. Dieses Layout is non-disclosed, entsprechend ist es unwahrscheinlich, dass man in CUDA Texturen als generische Buffer binden kann.

net werden soll und was nicht. In diesem Fall wird `glDrawElementsInstancedBaseVertex` für den OpenGL- Draw Call genutzt.

Ist kein Index Buffer gesetzt, werden die Vertices über `glDrawArraysInstanced` „direkt“ gezeichnet, also wird z.B. angenommen, dass drei hintereinander liegende Vertices ein Dreick bilden.

Um bei Ping Pong- Buffers vor dem Draw Call sicher zu gehen, dass der aktive Buffer gebunden ist, werden in der `draw()`-Routine, falls eine Flag anzeigt, dass Ping Pong Buffers unter den Attribute Buffers sind, diese neu ans VBO gebunden, falls der aktive Buffer seit dem letzten Draw call gewechselt hat.

3.13.3 InstancedGeometry

Dies Klasse stellt eine reine „Dummy-Geometrie“ dar. in Ihrer `draw()`-Routine teilt sie ihrem zugehörigen `InstanceManager` nur mit, dass diese Instanz beim nächsten Instanced Rendering-Draw Call ebenfalls gezeichnet werden will. Für Details sei auf Abschnitt 4.1.8.3 verwiesen.

3.14 Das MPP-Paket

Dieses Paket stellt dem *Flewnit*-System alle Funktionalität zur Verfügung, die mit GPU-Programmen zu tun hat, außer der OpenGL-Kontext-Erstellung, welche vom `WindowManager` übernommen wird.

3.14.1 MPP und ParallelComputeManager

Das `MPP`, kurz für „Massively Parallel Program“, ist die Basis-Klasse für Shader und `CLProgram`.

```
1 class MPP
2 : public SimulationObject43
3 {
4     FLEWNIT_BASIC_OBJECT DECLARATIONS;
5 public:
6     MPP(String name, SimulationDomain sd);
7     virtual ~MPP();
8     virtual void build()=0;
9 protected:
10    virtual void setupTemplateContext(TemplateContextMap& contextMap)=0;
11    virtual void validate()throw(SimulatorException)=0;
12    //for later debugging of the final code of a stage:
13    void writeToDisk(String sourceCode, Path where);
14};
```

⁴³SimulationObject ist schlicht eine Klasse, die Namen und Simulations-Domäne speichert, damit diese nicht überall neu definiert und Getter implementiert werden müssen.

Die `ParallelComputeManager`- Singleton-Klasse erstellt einen OpenCL-Kontext, der für CL/GL-Interoperabilität mit dem (zuvor vom `WindowManager` erstellten) OpenGL-Kontext assoziiert wird. Dann werden über OpenCL-Queries die Members einer `ParallelComputeDeviceInfo`-Instanz gesetzt, so dass System-global alle relevanten/interessanten Hardware Features bequem verfügbar sind.

Die Klasse besitzt die OpenCL-C++-Objekte, die mit dem Kontext assoziiert sind: Den `cl::Context` selbst, das assoziierte `cl::Device` (also für gewöhnlich die Repräsentation der GPU des Rechners), und die assoziierte `cl::CommandQueue`, in die die einzelnen Kernel Launches, Buffer/Image Reads/Writes/Copies, Synchronisations-Punkte etc. eingefügt werden, und stellt diese über Getter-Funktionen dem System zur Verfügung.

Der OpenCL-C++-Wrapper implementiert Error-Handling inklusive dem Werfen von `cl::Exceptions`. Bei der OpenGL-C-API muss man Errors aber nach wie vor explizit abfragen. Dieses - mit zusätzlich formatiertem Konsolen-Output - übernimmt `void checkCLGLErrors()` (checkt auch den letzten CL-Error, einfach zur Vollständigkeit, auch wenn es nicht nötig ist). Um bei der Entwicklung des Systems sofort auf OpenGL-Errors aufmerksam gemacht zu werden, ist folgendes Makro definiert:

```

1 #ifdef _DEBUG
2     //Macro to permanently check for errors in debug mode
3 # define GUARD(expression) \
4     expression; \
5     ParallelComputeManager::getInstancePtr()->checkCLGLErrors()
6 #else
7 # define GUARD(expression) expression
8 #endif

```

Sämtliche OpenGL-Calls sind irgendwie durch dieses Makro eingeschlossen. Der etwaige Performance-Overhead spielt durch bedingte Kompilierung im Release-Build keine Rolle.

Da auf die Klasse sehr häufig zugegriffen wird, ist das Shortcut-Makro `#define PARA_COMP_MANAGER ParallelComputeManager::getInstancePtr()` für bequemere Referenzierung der Singleton-Instanz im Code definiert.

`ParallelComputeManager` besitzt sämtliche `MPP`-Instanzen, die sich bei ihr über ihren Konstruktor automatisch registrieren, und ist für deren Löschung beim Engine-Reset zuständig.

Alle `BufferInterface`-Instanzen, die sowohl eine OpenCL- als auch eine OpenGL- Repräsentation haben, registrieren sich automatisch bei dieser Instanz, so dass die Akquisition von Shared Buffers für einen bestimmten Kontext ganz einfach geschehen kann:

```

1 void ParallelComputeManager::acquireSharedBuffersForCompute()
2 {
3     //skip if not necessary;
4     if(computeIsInControl()) return;
5     //force to wait for all GL commands to complete
6     barrierGraphics();44
7     mLastCLError = mCommandQueue.enqueueAcquireGLObjects(& ←
8         mRegisteredCLGLSharedBuffers);
9     mCLhasAcquiredSharedObjects = true;
}

```

`void acquireSharedBuffersForGraphics()` ist ähnlich implementiert.

Ob Buffer-Operationen mit dem Host-Code synchronisiert werden soll (also ob API-Calls wie `cl::CommandQueue::enqueueWriteImage(..)` bis zum Beenden der Buffer-Operation nicht zurückkehren sollen) oder nicht, lässt sich global über `void setBlockAfterEnqueue(cl_bool val)` und `cl_bool getBlockAfterEnqueue()const` setzen bzw abfragen (vgl. Listing 9 ff.). Letztendlich hat sich jedoch herausgestellt, dass Blocking bei Write-Operationen nicht nötig ist, weil ich (zumindest bei Kernel Launches) intensiv von `cl::Events` Gebrauch mache, auf von denen man Listen als Parameter an `cl::CommandQueue::enqueueXY()` übergeben kann, und dieser Command wird frühstens dann ausgeführt, wenn das alle Commands, die mit der Event-Liste assoziiert sind, vollständig ausgeführt sind. Blocking von Buffer-Operationen ist also nur dann nötig, wenn man nach einem Read-Back sofort im Host-Code die ausgelesenen Werte benötigt. Ich hätte mich beim Design noch mehr auf den Event-Waiting-Mechanismus fokussieren sollen, der zur Zeit vom `BufferInterface` noch nicht unterstützt ist. Ich wollte das `BufferInterface` so schlicht wie möglich halten. Momentan kommt deshalb die Synchronisation von Buffer-Operationen etwas holprig daher; Man kann den Design Flaw umgehen, indem man vor einer Buffer-Operation `PARA_COMP_MANAGER->getCommandQueue().enqueueWaitForEvents(eventVector)` aufruft, ein Refactoring ist dennoch in Planung.

3.14.2 Shader und ShaderManager

Die `Shader`-Klasse wrappt erwartungsgemäß die OpenGL-Shader-Funktionalität, stellt außerdem Convenience Functions zum Setzen von Uniform Variablen bereit, deren Kontrollfluss z.T. von den `ShaderFeaturesLocal` und den `ShaderFeaturesGlobal` abhängt. Das `MPP`-Interfaces ist implementiert, jedoch noch ableitbar. Nur die neue Routine `virtual void use(SubObject* so)throw(SimulatorException)=0;` ist rein abstrakt.

Inwiefern das Shader-Interface und seine Implementatioen geschickt

⁴⁴barrierGraphics() ist schlicht ein Wrapper für GUARD(glFinish()), so wie barrierCompute() als mCommandQueue.enqueueBarrier() implementiert ist.

und generisch gewählt sind oder nicht, konnte leider empirisch noch nicht umfassend ermittelt werden, da bisher der Shader-Template-Code aller abgeleiteten Shader-Klassen auf eine einzige Sammlung von Dateien zurück gehen, nämlich der mit dem suggestiven Namen „GenericLightingUberShader“; diese realisiert sowohl das klassische visuelle Rendering mit den in Abschnitt 4.1.8 vorgestellten Features, als auch die Shadow Map Generation und die Debug-Draw-Funktionalität.

Die Wahre Probe für die `Shader`-Klasse wird jedoch die, wenn eine Datei-Sammlung und mit ihr eine (Multi-Pass)-Shaderstruktur, die nicht viel mit dem „GenericLightingUbershader“ gemeinsam hat, zur Code-Basis wird. Inwiefern die Convenience Functions dann noch sinnvoll sind, (nicht-virtuell) in der Shader-Oberklasse definiert zu sein, vermag ich schwer abzuschätzen. Ob trotz der signifikanten Unterschiede diese Klasse gut abstrahiert ist und so allgemein nutzbar bleibt, wird sich erst dann herausstellen, wenn die `ParticleLiquidDrawStage` (s. Abb. 4) implementiert ist. Aus Zeitmangel konnte ich das Fluid noch nicht angemessen visualisieren, das `LiquidVisualMat` ist momentan noch – entgegen dem Klassendiagramm in Abb. 3 – „hacky“ von `DegugDrawVisualMat` abgeleitet.

In Anbetracht eines solchen nicht unerheblichen Vollständigkeits-Mangels bei der Implementierung sei hier explizit auf Abschnitt 3.15 verwiesen, der den Status der Implementierung zur Zeit der Abgabe der Ausarbeitung skizziert.

Die `ShaderManager`-Singleton-Klasse hat zum Ziel, so viele visuelle Rendering-Techniken und -Effekte in beliebigen (sinnvoller) sinnvollen Permutationen zu ermöglichen. Um nicht hunderte Shader von Hand anlegen zu müssen, die vor Boilerplate-Code überlaufen, wird wie gesagt die `Grantlee`-Template-Engine verwendet, um aus relativ wenigen Dateien angepasste Shader zu generieren.

Sie ist dafür zuständig, immer wenn sich das „Rendering Szenario“ ändert, also eine neue `LightinSimStageBase` im `LightingSimulator` aktiv wird, anhand von diversen Parametern allen `VisualMaterials` einen zum Material, dem Szenario und den globalen Shader-Features kompatiblen `Shader` zuzuweisen. Dieser wird entweder zur Laufzeit per `Grantlee` generiert, oder wenn schon ein Shader mit benötigten Features existiert, aus einer Hash-Map beschafft, wo der Hash-Wert der Shader-Features als Schlüssel dient. Das Sequenz-Diagramm in Abb. 6 skizziert den Ablauf.

Außerdem lassen sich einige Features abfragen, die ansonsten nur indirekt in den Feature-Structures encoded sind, wie z.B. ob die aktuelle Stage einen Framebuffer braucht oder nicht.

Die Parameter sind:

ShaderFeaturesGlobal Parameter, die global per Config gesetzt wurden und ohne Reset der Engine nicht mehr zu ändern sind, und aufgrund

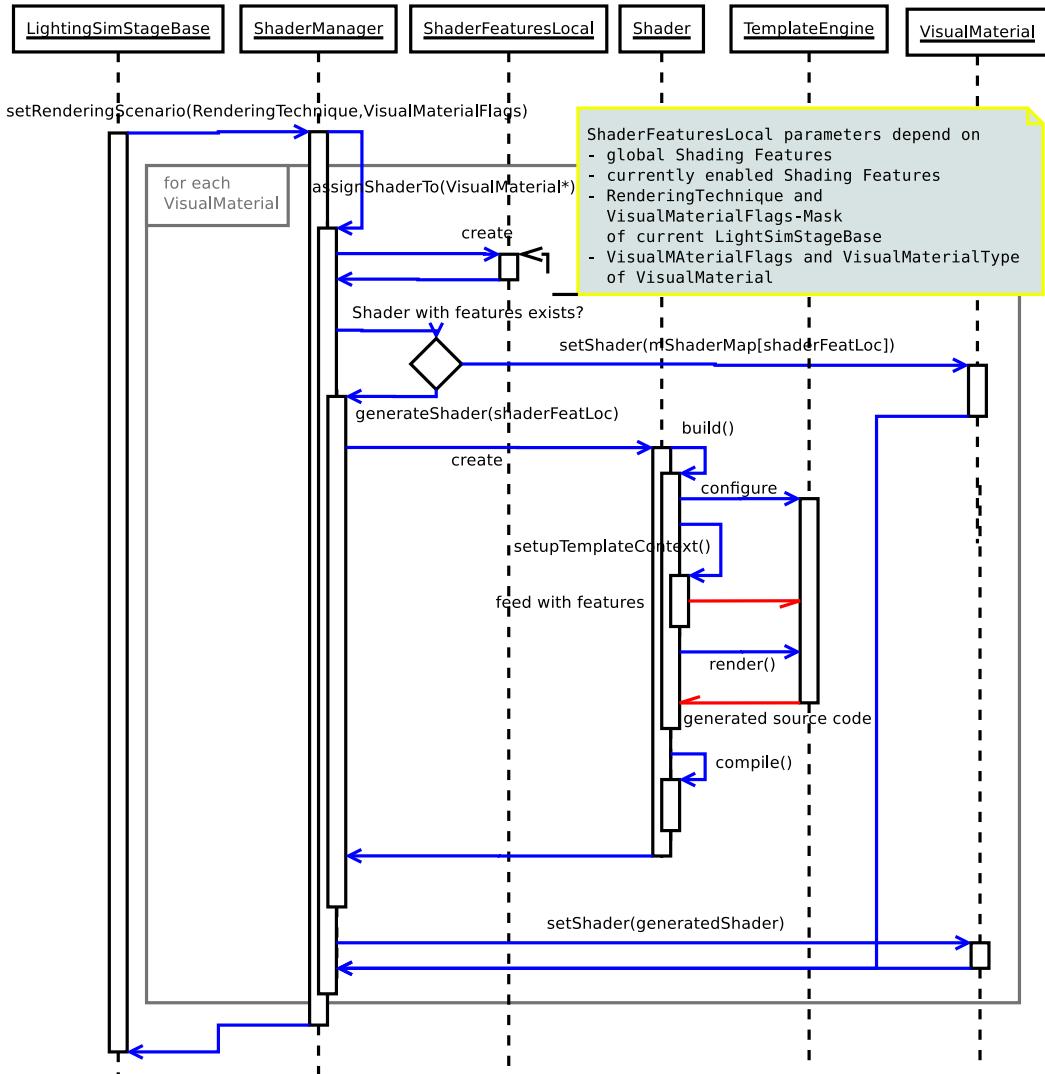


Abbildung 6: Sequenz-Diagramm der Shader-Generierung

von Buffergrößen, Kompatibilitäten und/oder der schlichten Konsistenz des visuellen Renderings sich zwischen den einzelnen Shadern nicht unterscheiden dürfen; dazu zählen:

- die maximale Anzahl an Lichtquellen
- die maximale Anzahl an Lichtquellen, die Schatten werfen (Shadow Caster),
- die maximale Anzahl an Instanzen, die pro Draw Call per Hardware Instancing gezeichnet werden kann
- falls man Deferred Rendering betreibt: Der Textur-Typ des G-Buffers (`Texture2D`, `Texture2DMultisample` etc.),

- falls man Deferred Rendering betreibt: Der MultiSample-Count des G-Buffers, falls er aus MultiSample-Texturen besteht
- das

```

1 enum LightSourcesLightingFeature
2 {
3     LIGHT_SOURCES_LIGHTING_FEATURE_NONE          =0,
4     LIGHT_SOURCES_LIGHTING_FEATURE_ONE_SPOT_LIGHT =1,
5     LIGHT_SOURCES_LIGHTING_FEATURE_ONE_POINT_LIGHT =2,
6     LIGHT_SOURCES_LIGHTING_FEATURE_ALL_POINT_LIGHTS =3,
7     LIGHT_SOURCES_LIGHTING_FEATURE_ALL_SPOT_LIGHTS =4,
8     LIGHT_SOURCES_LIGHTING_FEATURE_ALL_POINT_OR_SPOT_LIGHTS =5,
9     __NUM_LIGHT_SOURCES_LIGHTING_FEATURES__        =6
10 }

```

Dieser Aufzählungstyp delegiert bei der Shader-Generierung per Template-Engine, welche Typen und welche Anzahl an Lichtquellen vom Shader unterstützt werden sollen.

- das

```

1 enum LightSourcesShadowFeature
2 {
3     LIGHT_SOURCES_SHADOW_FEATURE_NONE          =0,
4     LIGHT_SOURCES_SHADOW_FEATURE_ONE_SPOT_LIGHT =1,
5     LIGHT_SOURCES_SHADOW_FEATURE_ONE_POINT_LIGHT =2,
6     LIGHT_SOURCES_SHADOW_FEATURE_ALL_SPOT_LIGHTS =3,
7     __NUM_LIGHT_SOURCES_SHADOW_FEATURES__        =4
8 }

```

Es gibt an, ob und wenn ja, welche Art Shadow Mapping vonstatten gehen soll.

- das

```

1 enum ShadowTechnique
2 {
3     SHADOW_TECHNIQUE_NONE      =0,
4     SHADOW_TECHNIQUE_DEFAULT   =1,
5     SHADOW_TECHNIQUE_PCFSS    =2,
6     __NUM_SHADOW_TECHNIQUES__ =3
7 }

```

, welches bestimmt, nach welchem Algorithmus Shadowmapping vonstatten gehen soll

diverse ShadingFeatures Diese Flags geben an, welche Shading-Features „aktiv“ sind: Die ShadingFeatures des letztendlich generierten Shaders hängen von den Shading Features des Materials selbst ab, aber auch von denen der aktuell aktiven `LightInSimStageBase` (eine `ShadowMapGenerationStage` hat z.B. gar keine „Shading“-Features) und von den aktuell global

aktivierten Features; Welche Features global aktiv sind, kann vom Benutzer bestimmt werden, hängt aber z.T. auch von der verwendeten OpenGL-Version ab; Tessellation kann z.B. erst ab OpenGL 4.0 verwendet werden.

```

1 enum ShadingFeatures
2 {
3     SHADING_FEATURE_NONE          =1<<0,
4     SHADING_FEATURE_DIRECT_LIGHTING =1<<1,
5     //global lighting via layered depth images or stuff... just a ←
6     //brainstroming, wont be implemented
7     SHADING_FEATURE_GLOBAL_LIGHTING =1<<2,
8     SHADING_FEATURE_DIFFUSE_TEXTURING =1<<3,
9     SHADING_FEATURE_DETAIL_TEXTURING =1<<4,
10    SHADING_FEATURE_NORMAL_MAPPING =1<<5,
11    SHADING_FEATURE_CUBE_MAPPING =1<<6,
12    SHADING_FEATURE_AMBIENT_OCCLUSION =1<<7,
13    SHADING_FEATURE_TESSELLATION =1<<8,
14    __NUM_SHADING_FEATURES__      =9
15 };

```

VisualMaterialType Der Typ bestimmt für gewöhnlich die Klasse des konkreten Shaders (erinnere: `Shader` ist eine abstrakte Basisklasse):

```

1 enum VisualMaterialType
2 {
3     VISUAL_MATERIAL_TYPE_NONE          =0,
4     VISUAL_MATERIAL_TYPE_DEFAULT_LIGHTING =1,
5     VISUAL_MATERIAL_TYPE_SKYDOME_RENDERING =2,
6     VISUAL_MATERIAL_TYPE_DEBUG_DRAW_ONLY =3, //just set a color ←
7     //value or something
8     VISUAL_MATERIAL_TYPE_GAS_RENDERING =4,
9     VISUAL_MATERIAL_TYPE_liquid_RENDERING =5,
10    __NUM_VISUAL_MATERIAL_TYPES__      =6
11 };

```

VisualMaterialFlags Diese Klasse trägt nicht nur zur Delegation der Shader Generierung bei, sie gibt auch bei einem `VisualMaterial` bestimmte Eigenschaften an, welche auf Kompatibilität mit den Flags der aktuellen `LightingSimStageBase` getestet werden können. Auf diese Weise können inkompatible Objekte maskiert werden. Der Konstruktor reicht aus für eine Idee:

```

1 VisualMaterialFlags(
2     bool castsShadows = true ,
3     bool isTransparent = false ,
4     bool isShadable = true ,
5     bool isDynamicCubeMapRenderable = true ,
6     bool isInstanced=false ,
7     bool isCustomMaterial=false );

```

3.14.2.1 Probleme

Es sei noch ein nicht so unwichtiges Wort verloren über den Versuch, viele Effekte und Techniken in beliebiger Permutation zu kombinieren: Da dank der Template-Engine nun die Möglichkeit besteht, in Vererbungs-Hierarchien zu denken und somit objektorientierte Ansätze ins Shader-Design einfließen zu lassen, zeigt sich umso krasser, wie schwer es ist, Effekte und Techniken sauber und modular zu kombinieren, wenn man obendrein noch maximale Performance haben will.

Viele Features in einem Shader bringen bei Integration neue Einschränkungen mit sich, die besonderer Behandlung bedürfen. Ein Beispiel ist die Kombination von Tessellation mit Shadow Mapping und/oder Layered Rendering: Hier tut sich das Problem auf, dass wir zur Shadowmap-Generierung aus Sicht der Lichtquelle rendern (die View-Matrix ist aus Position und Richtung der *Lichtquelle* konstruiert), wir aber für Level-Of-Detail-Berechnungen zur Bestimmung des Tessellation-Levels Vertex-Positionen im *Betrachter*-Viewspace benötigen! Dies hat den Grund, dass das LOD eines jeden Patches sowohl in der Shadow Map als auch im finalen Rendering gleich sein müssen! Wenn dies nämlich nicht der Fall ist, ist die Geometrie, die dem Shadow-Map-Lookup zugrunde liegt, eine andere als die des finalen Renderings; somit werden sprichwörtlich Äpfel mit Birnen verglichen, und es kann zu Artefakten kommen, z.B. Selbstverschattung, wo keine sein sollte (wenn durch Tessellation z.B. eine Furche generiert wird und die Lichtquelle weiter weg von der Geometrie ist als die Betrachter-Kamera, s. Abb. 7).

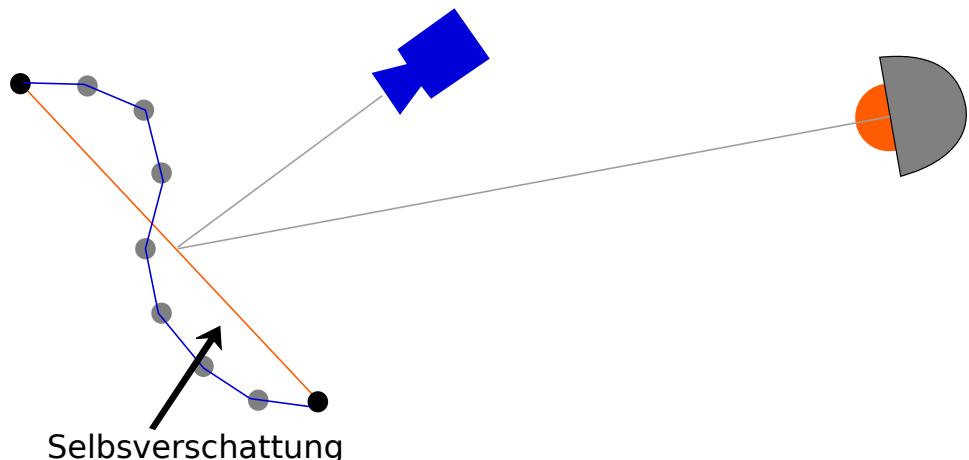


Abbildung 7: Artefakte bei naiver Tessellation-Level-Berechnung mit Camera-Viewspace bei der Shadow Map Generation

Um diese Artefakte zu vermeiden, muss für die LOD-Berechnung im Tessellation Control Shader *immer* die Viewspace-Position der Vertices aus

Sicht der Betrachter-Kamera zur Verfügung stehen. Bei der Shadow-Map-Generierung und/oder bei Layered Rendering (in Cube Maps und/oder Texture Arrays) werden aber für alles weitere, wie Displacement Mapping im Tessellation Evaluation Shader oder Beleuchtung im Fragment Shader, Viewspace-Positionen aus *Lichtquellen*-Sicht bzw *Worldspace*-Positionen durch die einzelnen Shader Stages geschleift.⁴⁵ Somit sind die klassischen Interface-Varying-Variable schon belegt. Man muss also eine neue Variable benutzen, um die Betrachter-Kamera-View Space-transformierte Vertex-Position vom Vertex Shader an den Tessellation Control Shader zu übergeben. Da die built-in-Variable `gl_Position` für OpenGL erst direkt vor der Fragment Shader-Stage für die Rasterisierung benötigt wird, kann man sie zuvor einfach zum Durchschleifen beliebiger Werte zweckentfremden. Somit ergibt sich folgender Vertex Shader Template-Code:

```

1 // step 2: calculate the gl_Position , if necessary
2 {%-if not layeredRendering and not SHADING_FEATURE_TESSELATION %}
3     //default case:
4     //we need projected transform for rasterization because no tessCtrl/←
        TessEval/geom shader follows the vertex shader;
5     gl_Position = modelViewProjectionMatrix * inPosition; /*default MVP ←
        transform*/
6 {%-endif %}
7 {%-if SHADING_FEATURE_TESSELATION %}
8     {%-if RENDERING_TECHNIQUE_SHADOWMAP_GENERATION %}
9         //write the "spectator view space position" to gl_Position for ←
            tessellation LOD calculations
10    gl_Position = spectatorCamViewMatrix * (modelMatrix * inPosition);
11 {%-else %}
12     {%- if worldSpaceTransform %}
13         //write the "own view space position" to gl_Position for ←
            tessellation LOD calculations;
14         //Hence, even for dynamic cubemap generation, a modelViewMatrix ←
            must be passed
15         //to the vertex shader for this scenario;
16         //The tessellation LOD calculation is invariant to cam rotation,
17         //hence only the translational part of the view matrix
18         //is relevant, and this part is equal for every six cube map faces;
19         gl_Position = modelViewMatrix * inPosition;
20     {%-endif %}
21 {%-endif %}
22 {%-endif %}
```

und für den Tessellation Control Shader:

```

1 //In order to determine the tessellation levels ,
2 //we need all three vertex positions in spectator view space
3 //of the triangle patch to which the current vertex belongs:
4 {%-if RENDERING_TECHNIQUE_SHADOWMAP_GENERATION or worldSpaceTransform %}
5     //every user varyings are either in spectator world space
```

⁴⁵bei Layered Rendering muss man bis zur Geomtry Shader Stage im World Space (nur mit der Model Matrix) rechnen, da im Geometry Shader jedes Layer eine andere Kamera, also eine andere View Matrix hat; Entsprechend kann man keine ModelView Matrix für frühzeitige View Space-Berechnungen akkumulieren!

```

6   //or in the light source- view- or world space;
7   //but we need the VIEW space vertex positions seen from the SPECTATOR ←
8   //camera for the tessellation level calculations!
9   //The vertex shader has written this value to gl_Position:
10  vec3 edgeStart = gl_in[ (gl_InvocationID + 1) % 3 ].gl_Position.xyz;
11  vec3 edgeEnd   = gl_in[ (gl_InvocationID + 2) % 3 ].gl_Position.xyz;
12  vec3 ownVert   = gl_in[ gl_InvocationID           ].gl_Position.xyz;
13  {% else %}
14  //no special case, we do just default rendering in spectator- view ←
15  //space; So grab the default position value:
16  vec3 edgeStart = input[ (gl_InvocationID + 1) % 3 ].position.xyz;
17  vec3 edgeEnd   = input[ (gl_InvocationID + 2) % 3 ].position.xyz;
18  vec3 ownVert   = input[ gl_InvocationID           ].position.xyz;
19  {% endif %}

```

Das Fazit ist: Modulares Shader-Design, welches effizienten Code produziert, scheint je nach Feature-Set schwer bis unmöglich. Das Nutzen von Vererbungsmechanismen wird durch die notwendigen Spezial-Behandlungen ebenfalls erschwert. Vor der Integration weiterer Features in diesen Über Shader muss die Struktur überarbeitet werden. Ich war zur Zeit des Schreibens des Shader-Codes weder mit dem Templating vertraut, noch habe ich alle Spezialfälle antizipieren können. Ich hoffe, dass mit neuen Einsichten und Erfahrungen Les- und Erweiterbarkeit vom Shader-Code durch ein Refactoring verbessert können, und dass sich geeignete Wege finden, trotz Problemen wie dem vorgestellten Komplexität durch Vererbung besser zu handhaben.

All diese Probleme betreffen nur den Shader-Code. Insbesondere die Vererbungsmechanismen lassen sich in OpenCL-Code-Templates reibungslos nutzen.

3.14.3 CLProgram und CLProgramManager

`CLProgram` und `CLKernel` wrappen die entsprechende OpenCL-Funktionalität: Das Programm repräsentiert kompletten ausführbaren Code, welches aus einem in OpenCL C geschriebenen Source Code kompiliert wurde. Ein Kernel ist dagegen eine von womöglich mehreren Funktionen innerhalb des Programms (durch das `__kernel`-Tag gekennzeichnet), welche vom Hostcode per `cl::CommandQueue::enqueueNDRangeKernel()` aufgerufen werden kann. Es besteht also eine $1 : n$ -Relation zwischen Programm und Kernel.

Die `CLProgramManager`-Singleton-Klasse gewährleistet einen Zugriff auf sämtliche OpenCL-Programme im System über ihren Namen; Dies kann nützlich sein, wenn ein Kernel auf die Beendigung eines anderen Kernels warten muss, dafür das Event, was mit der letzten Ausführung assoziiert ist, benötigt, aber kein Handle auf die entsprechenden `CLProgram`s vorhanden ist.

`CLProgramManager` besitzt einen `IntermediateResultBuffersManager`, der Buffers bereit stellt, die sich verschiedenen Kernels für ihre Zwischen-Ergebnisse teilen können. Somit wird GPU-Speicher gespart. Jedes `CLProgram`, welches Kernels hat, die Zwischenergebnisse benötigen, kann eine bestimmte

Menge und Größe an Buffers requesten. Es ist dann sichergestellt, dass der Buffer mit entsprechendem Index nach Allokation (geschieht nach Erschaffung sämtlicher `CLProgram`-Instanzen) mindestens die gewünschte Größe hat.

3.14.3.1 CLKernelArguments

Der Umgang mit Kernel-Argumenten, also Parametern, die an die Kernel-Funktionen übergeben werden, wird durch die Klasse `CLKernelArguments` und seine Members deutlich erleichtert: In OpenCL kann man Werte an Kernels nur über ihren Index in der formalen Parameter-Liste übergeben. Diese holprige Methode ist nicht nur im Code wenig aussagekräftig, sie birgt auch die Gefahr, dass wenn Kernel-Signaturen sich ändern, im Host-Code genutzte Indices invalid werden. Wenn diese auch noch verteilt im System auftauchen, wird der Aufwand und die Fehlerwahrscheinlichkeit noch größer, seinen Host-Code anzupassen. Aus diesem Grund erstellt jedes konkrete `CLProgram` seine eigenen Kernels, und erstellt auch eine Liste an Default-Parametern für jeden Kernel. Auf diese Parameter kann nun bequem sowohl per Index als auch über den Namen des Arguments zugegriffen werden. Typsicherheit ist ebenfalls durch die Template-Klasse `CLValueKernelArgument` gewährleistet, wo bei einem falschen Cast eine Exception geworfen wird.

Ferner behebt das `CLBufferKernelArgument` das bereits erwähnte Binding-Problem beim Toggle von PingPong-Buffers.

Ein detaillierteres Klassendiagramm zu den OpenCL-relevanten Klassen, welches zusätzlich die Call-Hierarchie des Build-Prozesses sowie einer Kernel-Invocation skizziert, zeigt Abb. 8.

3.15 Status der Implementierung am Ende der BA

Es sollten von Anfang an bei der Planung der *Unified Engine* viele Aspekte berücksichtigt werden, damit von vornherein ein flexibles und mächtiges System konzipiert werden konnte. Das Potential zur Weiterentwicklung war wichtiger als eine „schnell gehackte Demo“. Es wurden einige Features realisiert, andere sind nahezu vollständig programmiert, aber noch nie ausgeführt worden⁴⁶. Widerum andere Features liegen als Algorithmen in Textdateien herum, und es ergab sich noch keine Gelegenheit, diese Algorithmen in Programm-Code zu transformieren. Wieder andere Features wurden nur ganz grob im Hinterkopf behalten, um die Wahrscheinlichkeit zu erhöhen, dass sie in ferner Zukunft nahtlos mit den anderen Features integrierbar sind.

Eine Liste dieser Features, mit ihrem entsprechenden Implementations-Status

⁴⁶Mein Programmierstil ist sehr eigenwillig: ich kann wochenlang programmieren, ohne einmal compilieren, geschweige denn ausführen zu können. Dieser Stil scheint mir bei einem solchen Projekt nicht völlig verkehrt, da bei der sukzessiven Entwicklung der Fokus auf den Feature-Reichtum dem eines konsistenten Gesamtsystems vorgezogen werden könnte.

ist in Tabelle 4 und 5 zu finden.

Abb 9 zeigt einen Screenshot aus *Flewnit* mit der prototypischen Scene. Sie zeigt die interaktive partikelbasierte Fluid-Simulation inklusive der Visualisierung des Uniform Grid, eine weitere Box mit `DebugDrawMaterial`, einen Würfel aus Objekten, die per Instancing gezeichnet werden, tesseliert und normal mapped sind, einen Sky Dome, und schließlich ein tesselliertes, normal und environment mapped Raptor-Modell. Alle Objekte sind mit mehreren Lichtquellen, sowohl Spot als auch Point Light, in einem Render-Pass beleuchtet. Ein Spot Light wirft per Shadow Mapping Schatten. Details werden im Verlauf des Abschnitts 4 erläutert.

Der größte Wermutstropfen bei der Implementierung ist sicherlich die Visualisierung des Fluids, die minimalistischer technisch nicht sein könnte. Die Mechanik der Fluidsimulation stand erst wenige Wochen vor der Abgabe (die ersten Monate waren rein der Framework-Programmierung und dem nicht-Fluid-bezogenen visuellen Rendering gewidmet), so dass für eine angemessene Visualisierung des Fluids im Rahmen dieser Arbeit leider keine Zeit mehr blieb. Ich wollte schließlich meiner „ganz oder gar nicht“-Linie in diesem System treu bleiben, und werde die Visualisierung ohne „Zwischenlösung“ direkt implementieren, sobald Zeit ist. Hierfür möchte ich die in „Screen Space Fluid Rendering with Curvature Flow“ ([vdLGS07]) vorgestellte Lösung nach-implementieren.

Weiterhin fehlen noch sämtliche erweiterten Features und „*a posteriori*“-Optimierungen (also Optimierungen, die nach der erfolgreichen ersten Implementierung programmiert sind) bei der mechanischen Fluid-Simulation. Der Kernel- und Host-Code für die partikel-basierten Rigid Bodies ist fast komplett programmiert, da der Beginn der Ausarbeitung aber sehr dringend wurde, habe ich der Versuchung widerstanden, diesen Code noch zu testen.

Weiterhin wollte ich „static triangle collision meshes“ in die ansonsten partikelbasierte mechanische Simulation integrieren, über den Entwurf eines Algorithmus zur Voxelisierung von Dreiecken ins Uniform Grid bin ich jedoch nicht hinaus gekommen (mehr dazu im Ausblick, Kapitel 7).

Ferner habe ich zwar einen Scene-Loader mithilfe von *Assimp* implementiert, es ist aber noch vieles hart-Codiert, so dass das Laden von beliebigen Assets noch nicht möglich ist. Mein Fokus lag auf der Realisierung von kombinierbaren Effekten, vorerst nicht auf künstlerischem Anspruch. Dennoch ist die Fertigstellung des Scene Loaders geplant.

Die fehlende GUI erschwert das Parameter-Tweaken, was zur Zeit noch per XML-Datei geschehen muss.

Für weitere Details im Umgang mit fehlenden Features und unvollständiger Implementation sei auf den Ausblick in Abschnitt 7 verwiesen.

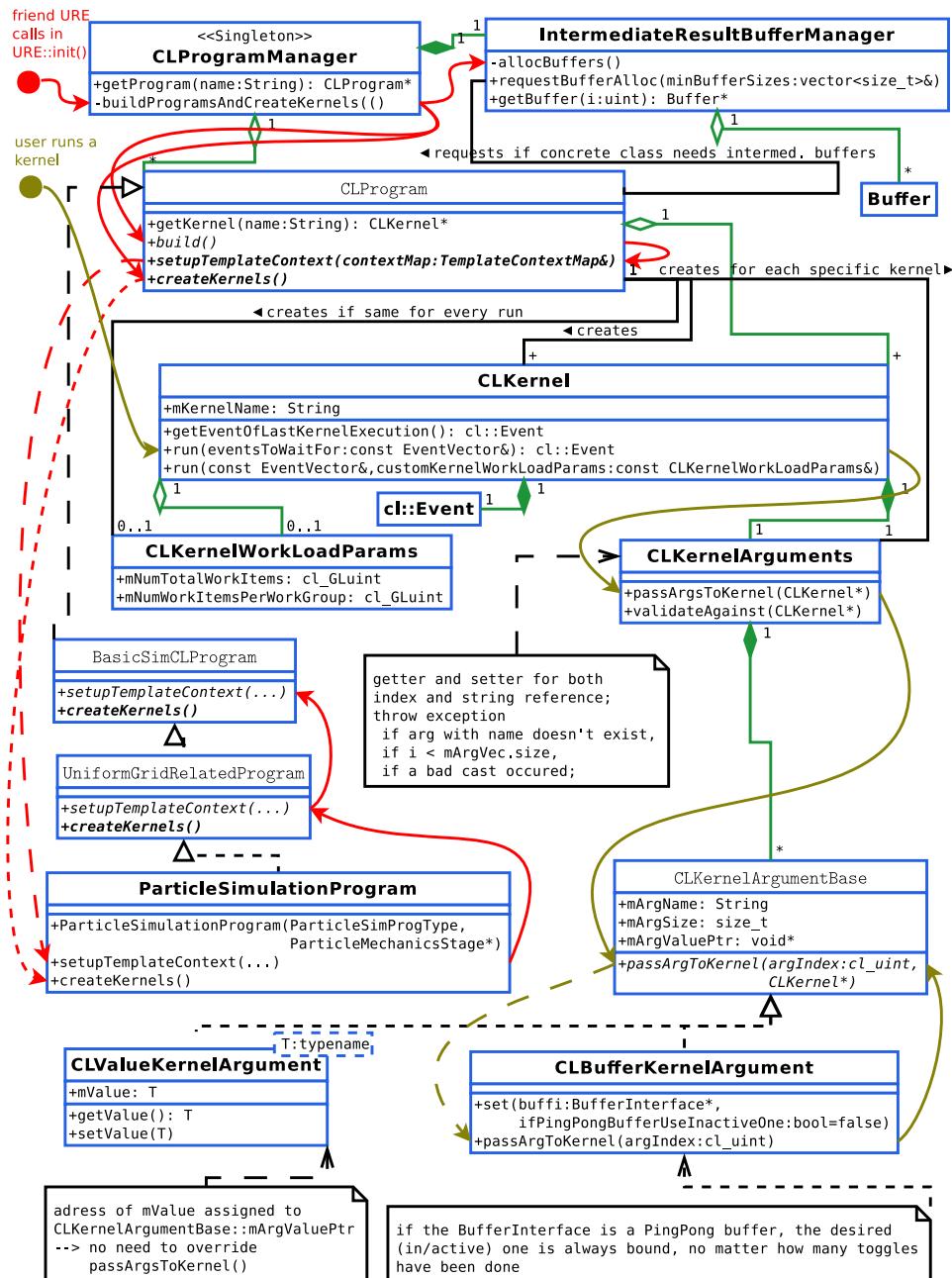


Abbildung 8: Klassendiagramm der OpenCL-relevanten Klassen inkl. Skizze der Call-Hierarchie des Build-Prozesses sowie dem Kernel-Launch

used OpenGL Features	Status	
Uniform Buffers	✓	
Instancing ^a	✓ ^a	
Tessellation ^b	✓	
Lighting	arbitrary light source amount in Shader point light lighting spot light lighting both point and spot light lighting	✓ ^c ✓ ✓ ✓
Deferred Rendering	Render to 2D Depth Texture Render to 2D Color Texture Render to MultiSample Color Texture G-Buffer Fill G-Buffer Shade G-Buffer MultiSample Shade Ambient Occlusion Post Processing ^d	✓ ○ ○ ○ ○ ○ ○ X X
Layered Rendering	Render to Cube Depth Texture Render to Cube Color Texture Render to Depth Texture Array Render to Cube Depth Texture Array	○ ○ ○ - ^e
Shadow Mapping Support	single spot light single point light ^f multiple spot light ^g multiple point light ^h	✓ ○ ○ -
Shadow Map Sampling	default PCFSS	✓ X
misc. Shading Features	arbitrary (useful) effect combination diffuse texturing detail texturing ⁱ Normal Mapping Environment Mapping Sky Dome transparency	✓ ✓ X ✓ ✓ ✓ X

Tabelle 4: Status der Implementation zum Zeitpunkt der Abgabe der Ausarbeitung (englisch wegen vielen Fachbegriffen) Legende:

✓ → implementiert; ○ → zu weiten Teilen programmiert, jedoch noch nicht integriert/ausgeführt/getestet; ○ → detailliert Konzipiert, jedoch nicht programmiert; X → im System langfristig konzipiert, jedoch nicht programmiert; - → Unterstützung nicht geplant

^aTransformation Matrix set provided per Uniform Buffers

^bproviding dynamic LOD, detail added via Displacement Mapping

^cAs long as they fit into an Uniform Buffer

^dParticipating Media, Depth Of Field etc.

^eonly supported in OpenGL 4

^frequires render to Cube Depth Texture 56

^grequires render to Depth Texture Array

^hrequires render to Cube Depth Texture Array

ⁱhigh frequency noise pattern to omit blurry appearance on close-up

		Status
mech. fluid simulation	basic efficient SPH mechanics ^a hardware dependent optimizations ^b algorithmic tricks & optimizations ^c multiple fluids particle Rigid Bodies static triangle collision mesh	✓ ✓ ○ ○ ○ ○
vis. fluid simulation ^d	debug draw curvature flow perlin noise foam rendering fresnel effect	✓ ✗ ✗ ✗ ✗

Tabelle 5: Status der Implementation zum Zeitpunkt der Abgabe der Ausarbeitung (Fortsetzung)

^asee section 4.2.1.5

^bsee section 4.2.1.6

^cintegrate density computation into force kernel, less bandwidth consuming integration scheme, see section 7

^dimplementation of [vdLGS07] is planned

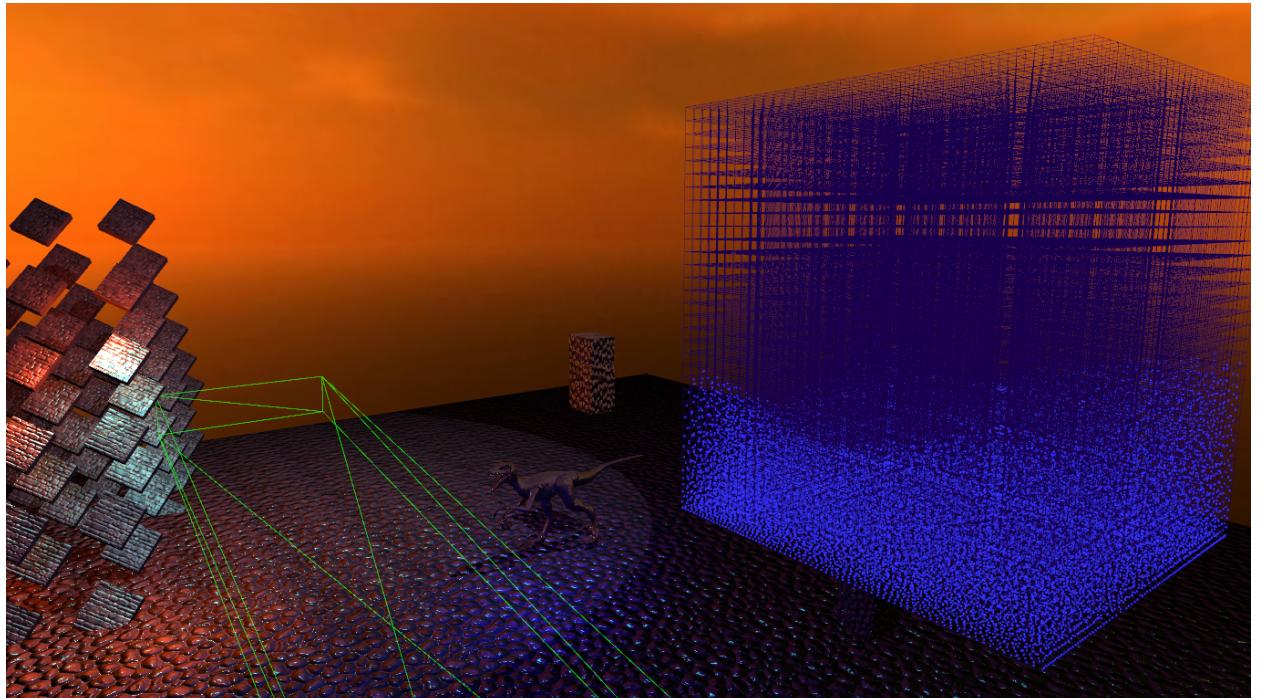


Abbildung 9: Überblick über die prototypische Szene

4 Simulation

4.1 Die visuelle Simulationsdomäne

Ein paar worte ueber die shading features, wie sie maskiert werden, SceneNodeVisitor etc..

4.1.1 Der LightingSimulator

Nochmal drauf hinweisen, dass Rendering etwas generisches in diesem Framework ist, und wir lieber von Lichtsimulation sprechen sollten, auch wenn es monetan nicht photrealistisch ist ;)

4.1.2 Die Lighting Simulation Pipeline Stages

baseclass etc... shadowmap gen stage, direct lighting stage, was noch in planung is etc..

in planung: deferred rendering G-Bufferfill, deferred rendering shade, div. post processing stages

4.1.3 VisualMaterial

4.1.4 Camera

4.1.5 LightSource und LightSourceManager

4.1.6 RenderTarget

4.1.7 ShaderManager

generiert mit grantlee, assigned an materials und verwaltet Shader , abhaengig von der aktuellen lighting stage, den registrierten Materials, der Erzeugten kontext, den vom user aktivierten rendering features etc pp

4.1.8 genutzte moderne OpenGL- Features

4.1.8.1 Hintergrund:Batching

referenz zu PCIe-Flaschenhals-stuff

PCIe-Bandbreite und -Latenz nicht überlasten durch immediate mode oder andere befehls-serien;

4.1.8.2 Uniform Buffers

auch von BufferInterface abstrahiert, vorteile auflisten, aber auch stolper-fallen: alignment etc) nutzen für transformationsmatrizen beim instancing und für beliebige lichtquekken

4.1.8.3 Instancing

InstanManager, InstangedGeometry vorstellen, konzept, wie es verwaltet wird, erklaeren; batching

Listing 4: Transformationsmatrizen-Uniforms im Vertex Shader

```
1  {% if instancedRendering %}  
2  
3      struct InstanceTransform  
4      {  
5          mat4 modelMatrix; //needed for layered rendering to be combined with←  
6              the several lightsource matrices  
7          mat4 modelViewMatrix; //needed in a non-layered context for ←  
8              calculation of view-space values for lighting calculations  
9          mat4 modelViewProjectionMatrix; //needed in a non-layered context for←  
10             gl_Position calculation  
11  
12         int uniqueInstanceID; //it is not guearanteed that for each "logic" ←  
13             instance, the gl_InstanceID stays the same for every draw call  
14                 //e.g. because of culling of "previous" ←  
15                 instances, the own gl_InstanceID will get ←  
16                 smaller  
17  
18     //no padding, because the offsets will be queried via ←  
19     glGetActiveUniformsiv (...)  
20 };  
21  
22     layout(shared) uniform InstanceTransformBuffer  
23     {  
24         InstanceTransform instanceTransforms[ {{numMaxInstancesRenderable}} ];  
25     };  
26  {% else %}  
27      uniform mat4 modelMatrix;  
28      uniform mat4 modelViewMatrix;  
29      uniform mat4 modelViewProjectionMatrix;  
30  {% endif %}
```

4.1.8.4 Hardware Tesselation

basics des GL4- hardware features erwaehnen fuer den geneigten leser, raptor-modell erwaehnen und seinen Aufbereitungsprozess, LOD, displacement mapping erlaeutern

4.1.9 Ablauf

4.1.10 Implementierte Effekte

diesen klumbatsch in form bringen, mit bildern anreichern etc pp

Zunächst zum Begriff "Mapping", der so oft auftaucht: Englisch "map- "Landkarte", "to mapäbbilden" Bedeutet in der Computergrafik meist die Abbildung eines Bildes auf eine Oberfläche nach einem bestimmten Algorithmus;

- Beleuchtung durch beliebig viele Punkt- und Spot-Lichtquellen (also Lichtquellen mit einem gerichteten Kegel, Scheinwerfer)
- Shadow Mapping: Erzeugen eines Bildes aus Tiefenwerten, anschließend vergleicht der Tiefenwerte aus Kamerasicht mit denen aus Lichtquellsicht (der "shadow map", Schattenkarte), pixel im finalen Bild gilt als verdeckt wenn Tiefenwert aus Kamerasicht größer als der entsprechende Pixel in der shadow map, unverdeckt wenn nicht;
- Normal Mapping: Verzerrung der Oberflächen-Normalen (Vektor senkrecht zur Oberfläche) um relief-artige Geometriedetails zu simulieren, ohne dass tatsächlich diese feine Geometrie in der virtuellen Szene existiert; Dies spart Rechenleistung und Speicher im Vergleich zu einer Szene, wo all dieses Detail tatsächlich in der Geometrie vorhanden wäre; Anschauliches Anwendungsbeispiel: Illusion der feinen Geometrie von Rauhfaserpete auf einem schlichten Quadrat; Nachteil: Die geometrische Illusion bricht bei flachen Betrachtungswinkeln ein, die Flachheit der eigentlichen, simplen Geometrie fällt dann auf; Die Information der verzerrten Normalen stammt ebenfalls aus einem Bild, der "normal map"; Diesmal werden die Pixelwerte jedoch nicht als Farben oder Tiefenwerte, sondern als Abweichung von den unverzerrten Normalen interpretiert (rot->x-Achse; grün->y-Achse; blau->z-Achse); Da im Computer alles nur Zahlen sind und Semantik erst durch unsere Verwendung und Wahrnehmung erlangen, und da die Graphikkarten so weit flexibel/programmierbar geworden sind, dass man als Programmierer Kontrolle über derartige "Überinterpretierung" hat, ist dies möglich;
- Environment mapping: Der Trick, perfekt spiegelnde Materialien vorgaukeln: Es wird in einer "Cube map" nachgeschaut, einer Sammlung von sechs Bildern, wo jedes Bild eine Würfelseite repräsentiert; Die Richtung der Normalen eines Pixels wird umgerechnet in eine Koordinate, mit der in der Cube Map nachgeschaut wird; Dieser Wert heißt dann in die Farbe des Pixels des finalen Bildes ein; Vorteil: Dinge wie lackierte Autotkarosserien lassen sich ganz gut vorgaukeln, mit recht geringem Rechenaufwand; Nachteil: Da für gewöhnlich nur in einem statischen Bild nachgeschaut wird, können dynamische Änderungen der Szene bei der "pseudo-Spiegelung" nicht erfasst werden; Ein Objekt, welches sich nahe eines Autos bewegt, bewegt sich in seiner Spiegelung nicht; Aus solchen Gründen sind in Cube maps oft nur sehr entfernte Dinge dargestellt: Horizont, Himmel, Wolken etc.; Diese Dinge ändern sich in der Realität ja nicht so schnell, daher fällt der Nachteil beim environment mapping unter dieser Einschränkung nicht mehr so drastisch auf; Der Hintergrund, die orangene Dämmerung, ist genau diese Cube map, die ich also sowohl für die Pseudo-Spiegelung als auch als "Füllmaterial" dort, wo ich keine Geometrie in der Szene habe, verwende;
- Tesselation: Wie bei Normal Mapping soll der wahrgenommene Detailgrad der Geometrie erhöht werden; Jedoch erzeugt die Tesselation „echte“ Geometrie, in Abhängigkeit von der Entfernung eines Objekts zur Betrachter-

Kamera; Somit wird dort Geometrie erzeugt, wo sie nötig für den Detailgrad des aktuellen Bildes ist, und dort eingespart, wo sie momentan unnötig ist; Diese Technik hat nicht die Nachteile des Normal mappings; Jedoch Ist durch die Reine Erzeugung von Geometrie noch nicht viel gewonnen; Sinn bekommt diese neue Geometrie wert dann, wenn sich auch wirklich mehr Datail mit ihr darstellen lässt; Erreicht wird dies durch eine sogenannte Displacement Map (frei Übersetzt "Verschiebungs-Karte", ein Bild, in dem Tiefenwerte der Hoch-detaillierten Geometrie gepeichert sind). Die neu erzeugte Geometrie wird also entlang der Normalen um den Betrag verschoben, wie in der Displacement Map eingetragen ist; Somit entsteht ein "tatsächliches Relief", im Gegensatz zum Vorgegaukelten Relief beim Normal Mapping; MEhr Details erspare ich dir, z.b. Warum man Normal Mapping trotzdem immer noch für die Beleuchtung braucht, trotz der Tessellation und dem Displacement mapping;

Anmerkung: Weil ich Tessellation so toll finde, habe ich mir das Velociraptor-3D-Modell aus dem Internet besorgt; Dieses hatte 5 Milloonen Dreiecke; Ich habe es mit einem Programm (was ich nicht selbst geschrieben habe, davon versteh ich leider noch viel zu wenig) herunterrechnen lassen, so dass ein vereinfachtes Modell mit etwa 11000 Dreicken entstand, also ein etwa zweitausend mal simpleres Modell. Mit einem anderen Programm habe ich dann die Geometrie des komplexen Modells auf die des simplen Modells projiziert, die detailgrad-bedingte Distanz zwischen den Geometrien in ein Bild geschrieben; Dieses Bild ist die Displacement map für die Tessellation zur Darstellung in meinem eigenen Programm; Somit kann ich nun den Dinosaurier beinahe so detailliert darstellen, wie er im Originalmodell vorliegt, jedoch mit viel höheren Bildwiederholungsraten;

4.2 Die mechanische Simulationsdomäne

blindtext blindtext

4.2.1 Fluidsimulation

blindtext blindtext

überlegen, ob ich aus Interesse nicht noch weiter in die Richtung recherchieren sollte, da ich nach meiner Implementierung erst so richtig beeindruckt von dem Verfahren war (ich habe im Internet noch keine Fluid-Demo gefunden, die ebenfalls SPH implementiert; ok., ich hab auch nicht gesucht ;), und gerne mehr über die Hintergründe verstehen würde... problem, wie immer: Zeitdruck ;(

4.2.1.1 Grundlagen

blindtext blindtext blindtext blindtext blindtext blindtext blindtext blindtext

4.2.1.1 .1 Die Navier-Stokes-Gleichungen

Herleitung, Erläuterung blindtext blindtext blindtext blindtext blindtext blindtext blindtext blindtext blindtext blindtext

4.2.1.1 .2 Grid-basierte vs. Partikelbasierte Simulation

blindtext blindtext

4.2.1.1 .2.1 Die zwei Sichtweisen: Lagrange vs. Euler

blindtext blindtext blindtext blindtext blindtext blindtext

4.2.1.1 .3 Smoothed Particle Hydrodynamics

ursprünglich aus astronomie blubb blubb

4.2.1.2 Verwandte Arbeiten

Referenzen auf Müller03, Thomas Steil, Goswami, GPU gems, Aufzeigen, was ich von wem übernommen habe, was ich selbst modifiziert habe aufgrund von etwaigen Fehlern in den Papers oder weil OpenCL es schlicht nicht zulässt;

4.2.1.3 Umfang

Abgrenzungf zwischen bisher funktionierenden Features, bisher programmierten, aber nicht integrierten und ungetesten Features und TODOs für die zukunft

4.2.1.4 Algorithmen

Verwaltung der Beschleunigungsstruktur ist der Löwenanteil, nicht die physik-simulation, die eher ein Dreizeiler ist;

4.2.1.4 .1 Work Efficient Parallel Prefix Sum

4.2.1.4 .2 Parallel Radix Sort und Stream Compaction

4.2.1.5 Ablauf

initialisierung, und beschreibung der einzelnen phasen...

4.2.1.6 Hardwarespezifische Optimierungen

5 Ergebnisse

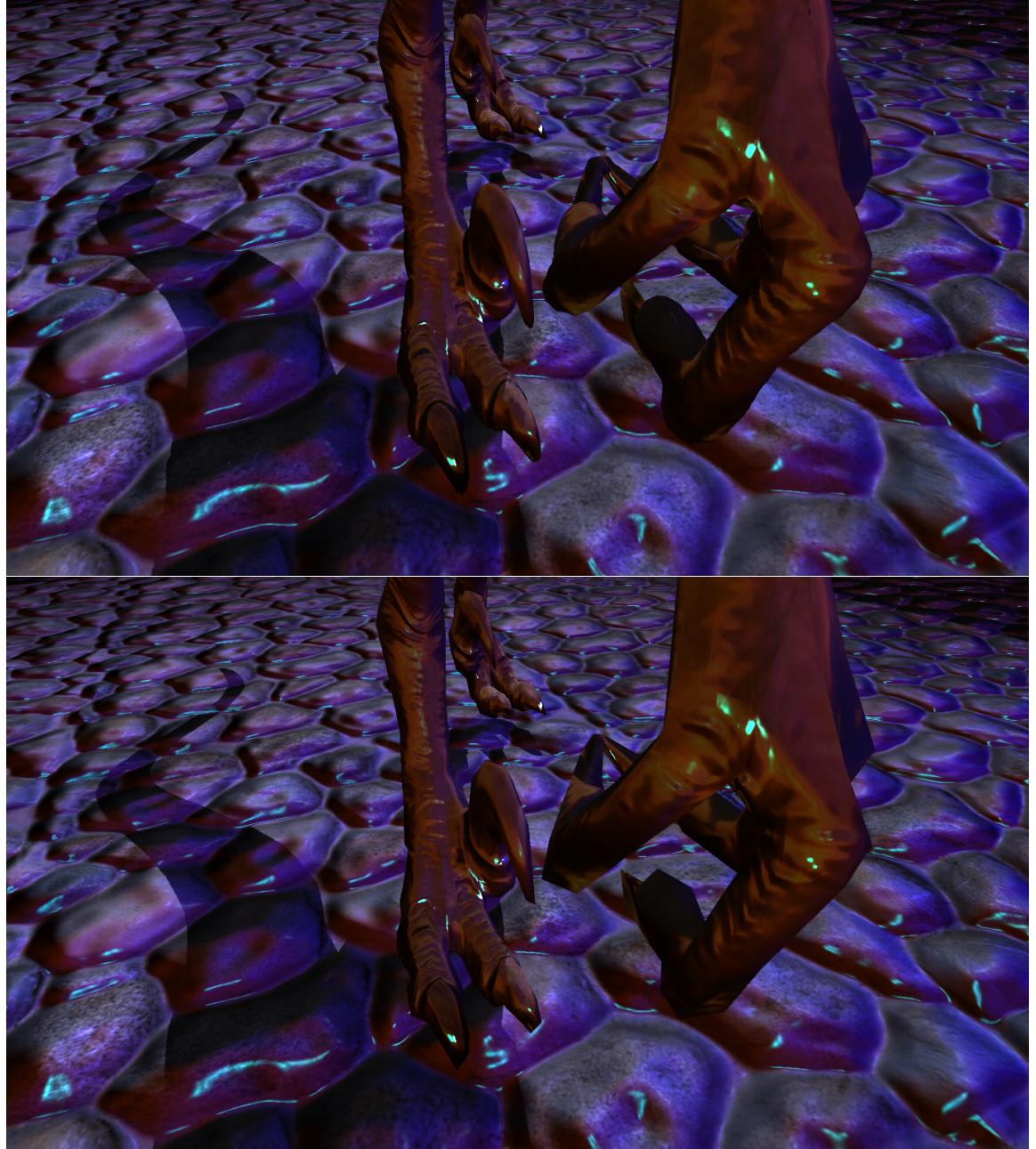


Abbildung 10: Die Extremitäten des Raptor-modells; Oben: tesseliert, texturiert, normal mapped, environment mapped, beleuchtet von fünf Lichtquellen , shadow mapped von einer Lichtquelle; Unten: wie oben, nur ohne Tessellation

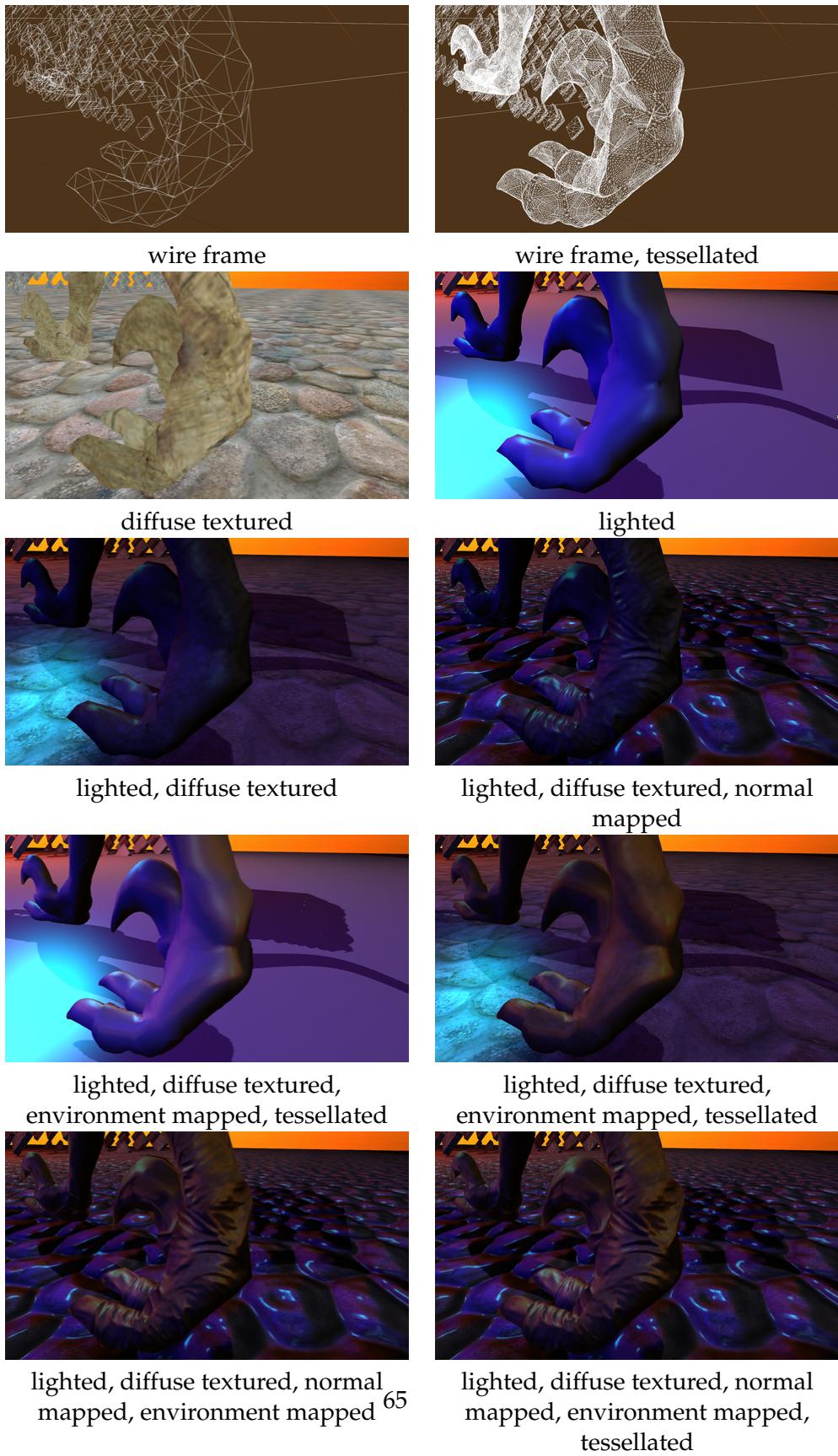
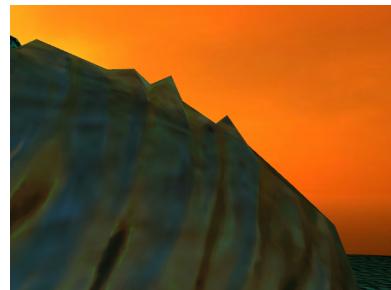


Abbildung 11: Füße des Raptor-Modells, mit verschiedenen Permutationen aktivierter Shader-Features

65
lighted, diffuse textured, normal mapped, environment mapped, tessellated



Auf die Entfernung reicht das Normal Mapping noch aus,
um Geometrie-Detail zu suggerieren ...



... hier jedoch nicht mehr



Abbildung 12: Fuß des Raptor-Modells, Gegenüberstellung des Detail-Grades mit und ohne Tessellation, Teil 1

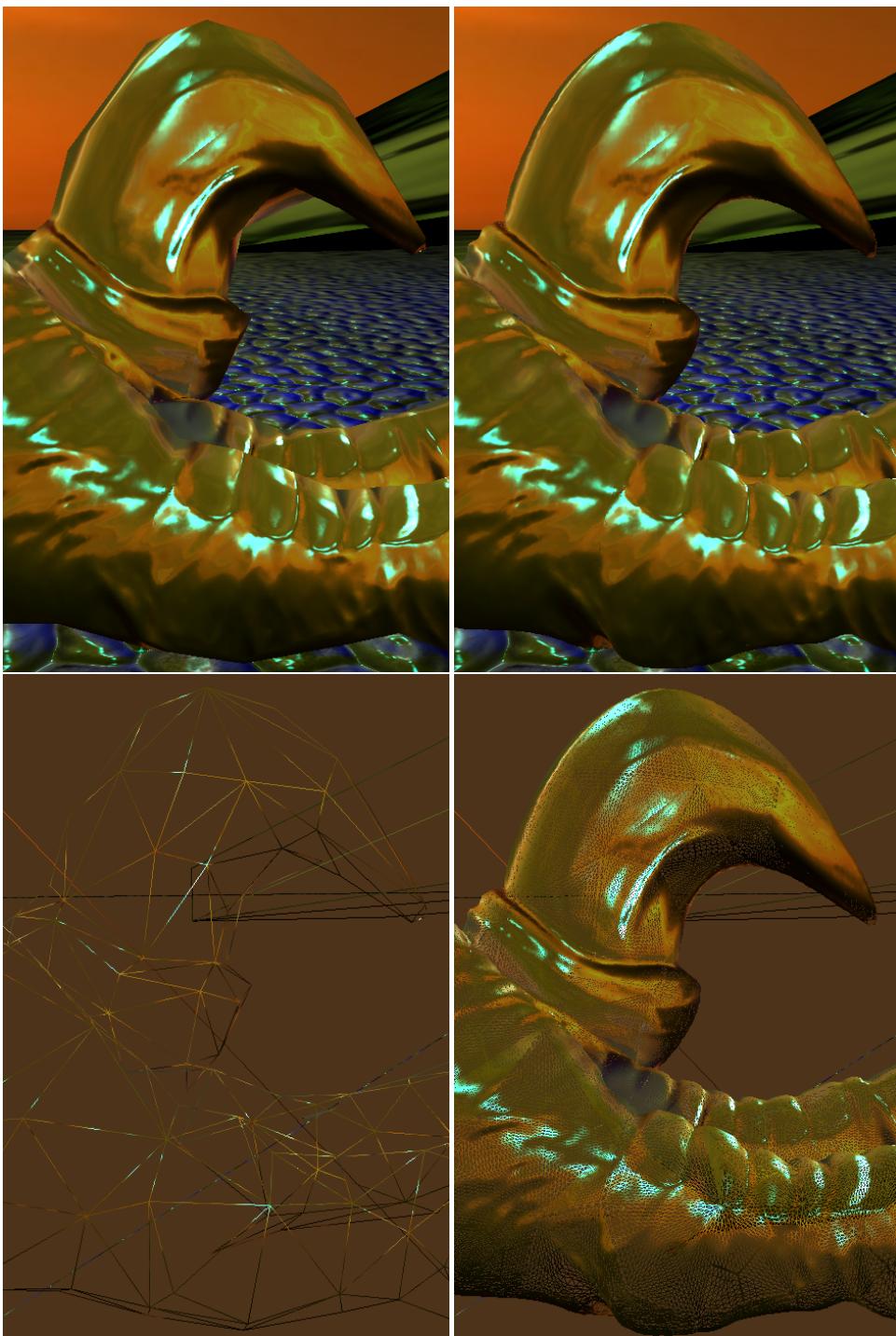


Abbildung 13: Fuß des Raptor-Modells, Gegenüberstellung des Detail-Grades mit und ohne Tessellation, Teil 2

6 Fazit

7 Ausblick

Zunächst möchte ich ergänzende Information zu dem Ansatz und der Motivation der „static triangle collision meshes“ liefern, die in Abschnitt 3.15 angedeutet wurden:

Mich hat es schon während der Recherche-Phase gestört, dass ich keine Information gefunden habe, ob und wenn ja, wie primitiv-basierte Rigid Body-Simulation – also eine Simulation, wo die Rigid Bodies aus Dreiecken oder komplexeren Primitiven aufgebaut sind und nicht aus Partikeln – auf der GPU effizient möglich ist. Ich habe nur Papers zu CPU/GPU-hybriden Ansätzen oder partikel-repräsentierten Rigid Bodies gefunden (letzteren Ansatz habe ich selber in meinem ungetesteten Code verwendet). Ich wollte zumindest für statische Dreiecks-Meshes testen, ob diese, da sich ihre räumliche Position und damit auch die Postion im Speicher nicht mehr ändert, effizient auf die GPU zu mappen sind. Hierfür habe ich einen Algorithmus entwickelt, der beschreibt, wie man das Mesh überhaupt erst einmal Artefakt-frei aufbereitet für die Integration in ein Uniform Grid (über Voxelisierung von Primitive-ID's, wobei der Umstand zu berücksichtigen ist, dass ein Vertex zu beliebig vielen Dreiecken gehören kann). Die nächste Herausforderung, wenn die Primitive-ID's wie Particle-ID's über das Uniform Grid verfügbar sind, ist die Datenstruktur für die Dreiecke und eine effiziente Kollisions-Berechnung zwischen Dreick und Partikel. Hierüber kann womöglich die Ray Tracing-Literatur mit ihren vielen Ansätzen zu effizienten Dreiecks-Repräsentationen und Strahl-Primitiv-Schnitttests Aufschluss geben; Beim Ray Tracing ist der effiziente Strahl-Primitiv-Test noch fundamentaler.

Der Mangel an Literatur ist kein gutes Omen für die Effizienz des Ansatzes. Dennoch finde ich es interessant, ihn weiter zu verfolgen, bis Klarheit über die effiziente Realisierbarkeit herrscht.

In Abschnitt 3.15 wurden die gröbsten Unvollständigkeiten der Implementierung bereits aufgezeigt. Dieses Kapitel fasst die fehlenden Features zu einer priorisierten „to do“-Liste zusammen:

1. Performance und Stabilität der Optimierung testen, welche Dichte-Berechnungen im selben Kernel ausführt wie die Kraft-Berechnungen (die Stabilität kann insofern leiden, dass bei diesem Ansatz immer die Dichte-Werte vom vorherigen Simulationsschritt verwendet werden müssen, welche somit veraltet sind)
2. Die partikelbasierte Rigid-Body-Simulation lauffähig machen
3. Ein anderes Integrations-Schema als das sehr Bandbreiten-lastige und Speicher-hungrige *Velocity Verlet*-Verfahren testen.

4. Das Fluid angemessen visualisieren nach [vdLGS07], aus Erkenntnissen hieraus womöglich die `Shader`-Klasse und den Shader-Template-Code des „GenericLightingUberShader“ refactorn
5. Die Funktionalität, mehrere Fluide zu simulieren und visualisieren, lauffähig machen
6. OpenCL-Kernels profilieren, Flaschenhälse finden, optimieren
7. Den `SceneLoader` fertig implementieren
8. Die Features zu Ende implementieren, für welche schon viel Code im System liegt: Deferred und Layered Rendering
9. Funktionalität zur Voxelisierung von Geometrie bereit stellen, so dass beliebige water tight meshes an der Rigid Body-Simulation teilnehmen können
10. Die „static triangle collision meshes“ implementieren
11. Eine GUI implementieren
12. In Hinblick auf Vision der Paddel-Simulation: *Wii Remote*-Anbindung inklusive *Wii MotionPlus*- Funktionalität⁴⁷ realisieren
13. weitere Features implementieren: Ambient Occlusion, Depth of Field etc.
14. „akustische Simulation“: testen, ob die generische Framework-Struktur sich auch auf diese Domäne übertragen lässt, und ob dies einen Mehrwert darstellt

Es gibt also genügend für die Zukunft zu tun. Es wird definitiv ein Langzeit-Projekt. Hoffentlich lässt der Zeitplan ein wenig Luft für die Weiterentwicklung.

⁴⁷http://en.wikipedia.org/wiki/Wii_Motion_Plus

[Sta99] [Sta03] [BMF07] [MCG03] [GSSP10] [vdLGS07] [Pes09] [Ste07]
 [Ebe04] [LG08] [HSO07]

A Listings zur Buffer-Abstraktion

Listing 5: BufferSemantics

```

1 enum BufferSemantics
2 {
3     // following semantics generally used in
4     //Generic Vertex Attribute Buffers and/or OpenCL (interop) Buffers
5     //\{
6     POSITION_SEMATRICS,
7     NORMAL_SEMATRICS,
8     TANGENT_SEMATRICS,
9     TEXCOORD_SEMATRICS,
10
11    VELOCITY_SEMATRICS,
12    MASS_SEMATRICS,
13    DENSITY_SEMATRICS,
14    PRESSURE_SEMATRICS,
15    FORCE_SEMATRICS,
16
17    Z_INDEX_SEMATRICS,
18    DIFFUSE_COLOR_SEMATRICS,
19    CUSTOM_SEMATRICS,
20
21    //we need that value to have static-length arrays holding VBO ←
22    //maintainance information
22    __NUM_VALID_GEOMETRY_ATTRIBUTE_SEMATRICS__,
23
24    //Semantic of the index buffer from a VBO used via glDrawElements()
25    INDEX_SEMATRICS,
26
27    //for a uniform buffer for matrices of instanced rendering
28    TRANSFORMATION_MATRICES_SEMATRICS,
29    LIGHT_SOURCE_BUFFER_SEMATRICS,
30    //\}
31
32    //following texture-only semantics; Texture can also have most of the ←
32    //above Semantics
33
34    DISPLACEMENT_SEMATRICS, //normal-depth or normal map
35    ENVMAP_SEMATRICS,
36    MATERIAL_ID_SEMATRICS,
37    PRIMITIVE_ID_SEMATRICS,
38    SHADOW_MAP_SEMATRICS,
39    AMBIENT_OCCLUSION_SEMATRICS, //attenuation value as result of AO ←
39    //calculations on a depth/position buffer
40    DETAIL_TEXTURE_SEMATRICS, //for stuff like terrain to hide low res ←
40    //decal texture
41    NOISE_SEMATRICS,
42    DEPTH_BUFFER_SEMATRICS, //e.g. for memory optimized ambient occlusion ←
42    //calculation
43    STENCIL_BUFFER_SEMATRICS,
44    INTERMEDIATE_RENDERING_SEMATRICS,
45    FINAL_RENDERING_SEMATRICS,
```

```

46
47 //actually real amount is one less than this value, but that doesn't matter
48 __NUM_TOTAL_SEMATRICS__,
49
50 //indicator for "empty" stuff, e.g. an empty Color Attachment slot in an FBO
51 INVALID_SEMATRICS
52 };

```

Listing 6: GLImageFormat-Definition für intern symmetrischeren Umgang mit den Type-/Layout-/Precision-/Normalization- Makros von OpenCL und OpenGL

```

1 struct GLImageFormat
2 {
3     // "internalformat" parameter of glTexImageXD; specifies
4     // - number of channels
5     // - number of bits per channel
6     // - data type of channel values (float, uint, int)
7     // - if (u)int values shall be normalized;
8     // All of this packed in one enum! Plus, this enum must comply with the
9     // channelOrder
10    // and the channelDataType below! So much invalid combinations are
11    // possible!
12    // Thus, the values of those enums are determined
13    // valid values: see the big table within the comment of
14    // BufferElementInfo: GL_RGBA32F .. GL_R8_SNORM;
15    GLint desiredInternalFormat;
16
17    // "format" parameter of glTexImageXD resp. cl_channel_order:
18    // texelInfo.numChannels == 1 --> GL_RED, 2 --> GL_RG, 4 --> GL_RGBA
19    GLenum channelOrder;
20
21    // "type" parameter of glTexImageXD, i.e cpu data layout;
22    // Compared to OpenCL, the if a (u)int- texel is normalized during
23    // texture lookup or not is
24    // determined via the "internalformat" param and not together with the
25    // CPU data type specifier;
26    // Therefore, only the following values are allowed;
27    // GL_FLOAT |GL_INT |GL_UNSIGNED_INT
28    // GL_HALF_FLOAT |GL_SHORT |GL_UNSIGNED_SHORT
29    // — |GL_BYTE |GL_UNSIGNED_BYTE
30    GLenum channelDataType;
31 };

```

Listing 7: BufferElementInfo

```

1 struct BufferElementInfo
2 {
3     /*
4      In this framework, it is strongly recommended to specify
5      the desired internal layout exactly, i.e. specify
6      1. number of channels: 1,2 or 4;
7          three channel types are not supported
8          by this framework due to alignment reasons;
9      2. internal data format on GPU memory: float,int, unsigned int
10         3. byte size per element: 8,16, or 32

```

```

11     4. flag , if unsigned integer data formats
12         should be normalized to float in [0..1]
13         resp. if signed integer data formats
14         should be normalized to float in [-1..1]
15
16     There is one minor drawback at the moment:
17     CPU data must have the same format as the data
18     which will reside on the GPU;
19     */
20
21     //will be mapped internally to:
22     // the "format" param of glTexImageXD:
23     // GL_RED, GL_RG, GL_RGB (forbidden), GL_RGBA resp.
24     // the cl::ImageFormat.image_channel_order:
25     // CL_R, CL_RG, CL_RGB (forbidden), CL_RGBA
26     int numChannels;
27     GPU_DataType internalGPU_DataType;
28     int bitsPerChannel; //8,16 or 32
29     bool normalizeIntegralValuesFlag;
30
31     //guard to tell the validate() function that this struct is unused
32     //(e.g. for lightsource uniform buffers, generic opencl buffers etc);
33     //is set to true in the default constructor,
34     //copied in the copy constructor and set to false in the value-passing ←
35     //consturctor
36     bool hasNoChanneledElements;
37
38     /*... Constructors, equality and assignment operators omitted ...*/
39
40     void validate() const throw (BufferException);
41 };

```

Listing 8: Mapping von BufferElementInfo-Members zu OpenGL/OpenCL-Makros, exemplarischer Ausschnitt

```

1  /*
2
3  bool normalize = elementInfo.normalizeIntegralValuesFlag;
4
5  switch(elementInfo.internalGPU_DataType)
6  {
7  case GPU_DATA_TYPE_UINT :
8      switch(elementInfo.bitsPerChannel)
9      {
10         case 8:
11             glImageFormat.channelDataType = GL_UNSIGNED_BYTE;
12             if(normalize) clImageFormat.image_channel_data_type = CL_UNORM_INT8;
13             else          clImageFormat.image_channel_data_type = CL_UNSIGNED_INT8;
14             switch(elementInfo.numChannels)
15             {
16                 case 1:
17                     elementType = TYPE_UINT8;
18                     if(normalize) glImageFormat.desiredInternalFormat = GL_R8;
19                     else          glImageFormat.desiredInternalFormat = GL_R8UI;
20                     break;
21                 case 2:
22                     elementType = TYPE_VEC2UI8;
23                     if(normalize) glImageFormat.desiredInternalFormat = GL_RG8;
24                     else          glImageFormat.desiredInternalFormat = GL_RG8UI;
25                     break;
26             }
27         }
28     }
29 }
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
589
590
591
592
593
594
595
596
597
598
599
599
600
601
602
603
604
605
606
607
608
609
609
610
611
612
613
614
615
616
617
618
619
619
620
621
622
623
624
625
626
627
628
629
629
630
631
632
633
634
635
636
637
637
638
639
639
640
641
642
643
644
645
645
646
647
647
648
649
649
650
651
652
653
653
654
655
655
656
656
657
657
658
658
659
659
660
660
661
661
662
662
663
663
664
664
665
665
666
666
667
667
668
668
669
669
670
670
671
671
672
672
673
673
674
674
675
675
676
676
677
677
678
678
679
679
680
680
681
681
682
682
683
683
684
684
685
685
686
686
687
687
688
688
689
689
690
690
691
691
692
692
693
693
694
694
695
695
696
696
697
697
698
698
699
699
700
700
701
701
702
702
703
703
704
704
705
705
706
706
707
707
708
708
709
709
710
710
711
711
712
712
713
713
714
714
715
715
716
716
717
717
718
718
719
719
720
720
721
721
722
722
723
723
724
724
725
725
726
726
727
727
728
728
729
729
730
730
731
731
732
732
733
733
734
734
735
735
736
736
737
737
738
738
739
739
740
740
741
741
742
742
743
743
744
744
745
745
746
746
747
747
748
748
749
749
750
750
751
751
752
752
753
753
754
754
755
755
756
756
757
757
758
758
759
759
760
760
761
761
762
762
763
763
764
764
765
765
766
766
767
767
768
768
769
769
770
770
771
771
772
772
773
773
774
774
775
775
776
776
777
777
778
778
779
779
780
780
781
781
782
782
783
783
784
784
785
785
786
786
787
787
788
788
789
789
790
790
791
791
792
792
793
793
794
794
795
795
796
796
797
797
798
798
799
799
800
800
801
801
802
802
803
803
804
804
805
805
806
806
807
807
808
808
809
809
810
810
811
811
812
812
813
813
814
814
815
815
816
816
817
817
818
818
819
819
820
820
821
821
822
822
823
823
824
824
825
825
826
826
827
827
828
828
829
829
830
830
831
831
832
832
833
833
834
834
835
835
836
836
837
837
838
838
839
839
840
840
841
841
842
842
843
843
844
844
845
845
846
846
847
847
848
848
849
849
850
850
851
851
852
852
853
853
854
854
855
855
856
856
857
857
858
858
859
859
860
860
861
861
862
862
863
863
864
864
865
865
866
866
867
867
868
868
869
869
870
870
871
871
872
872
873
873
874
874
875
875
876
876
877
877
878
878
879
879
880
880
881
881
882
882
883
883
884
884
885
885
886
886
887
887
888
888
889
889
890
890
891
891
892
892
893
893
894
894
895
895
896
896
897
897
898
898
899
899
900
900
901
901
902
902
903
903
904
904
905
905
906
906
907
907
908
908
909
909
910
910
911
911
912
912
913
913
914
914
915
915
916
916
917
917
918
918
919
919
920
920
921
921
922
922
923
923
924
924
925
925
926
926
927
927
928
928
929
929
930
930
931
931
932
932
933
933
934
934
935
935
936
936
937
937
938
938
939
939
940
940
941
941
942
942
943
943
944
944
945
945
946
946
947
947
948
948
949
949
950
950
951
951
952
952
953
953
954
954
955
955
956
956
957
957
958
958
959
959
960
960
961
961
962
962
963
963
964
964
965
965
966
966
967
967
968
968
969
969
970
970
971
971
972
972
973
973
974
974
975
975
976
976
977
977
978
978
979
979
980
980
981
981
982
982
983
983
984
984
985
985
986
986
987
987
988
988
989
989
990
990
991
991
992
992
993
993
994
994
995
995
996
996
997
997
998
998
999
999
1000
1000
1001
1001
1002
1002
1003
1003
1004
1004
1005
1005
1006
1006
1007
1007
1008
1008
1009
1009
1010
1010
1011
1011
1012
1012
1013
1013
1014
1014
1015
1015
1016
1016
1017
1017
1018
1018
1019
1019
1020
1020
1021
1021
1022
1022
1023
1023
1024
1024
1025
1025
1026
1026
1027
1027
1028
1028
1029
1029
1030
1030
1031
1031
1032
1032
1033
1033
1034
1034
1035
1035
1036
1036
1037
1037
1038
1038
1039
1039
1040
1040
1041
1041
1042
1042
1043
1043
1044
1044
1045
1045
1046
1046
1047
1047
1048
1048
1049
1049
1050
1050
1051
1051
1052
1052
1053
1053
1054
1054
1055
1055
1056
1056
1057
1057
1058
1058
1059
1059
1060
1060
1061
1061
1062
1062
1063
1063
1064
1064
1065
1065
1066
1066
1067
1067
1068
1068
1069
1069
1070
1070
1071
1071
1072
1072
1073
1073
1074
1074
1075
1075
1076
1076
1077
1077
1078
1078
1079
1079
1080
1080
1081
1081
1082
1082
1083
1083
1084
1084
1085
1085
1086
1086
1087
1087
1088
1088
1089
1089
1090
1090
1091
1091
1092
1092
1093
1093
1094
1094
1095
1095
1096
1096
1097
1097
1098
1098
1099
1099
1100
1100
1101
1101
1102
1102
1103
1103
1104
1104
1105
1105
1106
1106
1107
1107
1108
1108
1109
1109
1110
1110
1111
1111
1112
1112
1113
1113
1114
1114
1115
1115
1116
1116
1117
1117
1118
1118
1119
1119
1120
1120
1121
1121
1122
1122
1123
1123
1124
1124
1125
1125
1126
1126
1127
1127
1128
1128
1129
1129
1130
1130
1131
1131
1132
1132
1133
1133
1134
1134
1135
1135
1136
1136
1137
1137
1138
1138
1139
1139
1140
1140
1141
1141
1142
1142
1143
1143
1144
1144
1145
1145
1146
1146
1147
1147
1148
1148
1149
1149
1150
1150
1151
1151
1152
1152
1153
1153
1154
1154
1155
1155
1156
1156
1157
1157
1158
1158
1159
1159
1160
1160
1161
1161
1162
1162
1163
1163
1164
1164
1165
1165
1166
1166
1167
1167
1168
1168
1169
1169
1170
1170
1171
1171
1172
1172
1173
1173
1174
1174
1175
1175
1176
1176
1177
1177
1178
1178
1179
1179
1180
1180
1181
1181
1182
1182
1183
1183
1184
1184
1185
1185
1186
1186
1187
1187
1188
1188
1189
1189
1190
1190
1191
1191
1192
1192
1193
1193
1194
1194
1195
1195
1196
1196
1197
1197
1198
1198
1199
1199
1200
1200
1201
1201
1202
1202
1203
1203
1204
1204
1205
1205
1206
1206
1207
1207
1208
1208
1209
1209
1210
1210
1211
1211
1212
1212
1213
1213
1214
1214
1215
1215
1216
1216
1217
1217
1218
1218
1219
1219
1220
1220
1221
1221
1222
1222
1223
1223
1224
1224
1225
1225
1226
1226
1227
1227
1228
1228
1229
1229
1230
1230
1231
1231
1232
1232
1233
1233
1234
1234
1235
1235
1236
1236
1237
1237
1238
1238
1239
1239
1240
1240
1241
1241
1242
1242
1243
1243
1244
1244
1245
1245
1246
1246
1247
1247
1248
1248
1249
1249
1250
1250
1251
1251
1252
1252
1253
1253
1254
1254
1255
1255
1256
1256
1257
1257
1258
1258
1259
1259
1260
1260
1261
1261
1262
1262
1263
1263
1264
1264
1265
1265
1266
1266
1267
1267
1268
1268
1269
1269
1270
1270
1271
1271
1272
1272
1273
1273
1274
1274
1275
1275
1276
1276
1277
1277
1278
1278
1279
1279
1280
1280
1281
1281
1282
1282
1283
1283
1284
1284
1285
1285
1286
1286
1287
1287
1288
1288
1289
1289
1290
1290
1291
1291
1292
1292
1293
1293
1294
1294
1295
1295
1296
1296
1297
1297
1298
1298
1299
1299
1300
1300
1301
1301
1302
1302
1303
1303
1304
1304
1305
1305
1306
1306
1307
1307
1308
1308
1309
1309
1310
1310
1311
1311
1312
1312
1313
1313
1314
1314
1315
1315
1316
1316
1317
1317
1318
1318
1319
1319
1320
1320
1321
1321
1322
1322
1323
1323
1324
1324
1325
1325
1326
1326
1327
1327
1328
1328
1329
1329
1330
1330
1331
1331
1332
1332
1333
1333
1334
1334
1335
1335
1336
1336
1337
1337
1338
1338
1339
1339
1340
1340
1341
1341
1342
1342
1343
1343
1344
1344
1345
1345
1346
1346
1347
1347
1348
1348
1349
1349
1350
1350
1351
1351
1352
1352
1353
1353
1354
1354
1355
1355
1356
1356
1357
1357
1358
1358
1359
1359
1360
1360
1361
1361
1362
1362
1363
1363
1364
1364
1365
1365
1366
1366
1367
1367
1368
1368
1369
1369
1370
1370
1371
1371
1372
1372
1373
1373
1374
1374
1375
1375
1376
1376
1377
1377
1378
1378
1379
1379
1380
1380
1381
1381
1382
1382
1383
1383
1384
1384
1385
1385
1386
1386
1387
1387
1388
1388
1389
1389
1390
1390
1391
1391
1392
1392
1393
1393
1394
1394
1395
1395
1396
1396
1397
1397
1398
1398
1399
1399
1400
1400
1401
1401
1402
1402
1403
1403
1404
1404
1405
1405
1406
1406
1407
1407
1408
1408
1409
1409
1410
1410
1411
1411
1412
1412
1413
1413
1414
1414
1415
1415
1416
1416
1417
1417
1418
1418
1419
1419
1420
1420
1421
1421
1422
1422
1423
1423
1424
1424
1425
1425
1426
1426
1427
1427
1428
1428
1429
1429
1430
1430
1431
1431
1432
1432
1433
1433
1434
1434
1435
1435
1436
1436
1437
1437
1438
1438
1439
1439
1440
1440
1441
1441
1442
1442
1443
1443
1444
1444
1445
1445
1446
1446
1447
1447
1448
1448
1449
1449
1450
1450
1451
1451
1452
1452
1453
1453
1454
1454
1455
1455
1456
1456
1457
1457
1458
1458
1459
1459
1460
1460
1461
1461
1462
1462
1463
1463
1464
1464
1465
1465
1466
1466
1467
1467
1468
1468
1469
1469
1470
1470
1471
1471
1472
1472
1473
1473
1474
1474
1475
1475
1476
1476
1477
1477
1478
1478
1479
1479
1480
1480
1481
1481
1482
1482
1483
1483
1484
1484
1485
1485
1486
1486
1487
1487
1488
1488
1489
1489
1490
1490
1491
1491
1492
1492
1493
1493
1494
1494
1495
1495
1496
1496
1497
1497
1498
1498
1499
1499
1500
1500
1501
1501
1502
1502
1503
1503
1504
1504
1505
1505
1506
1506
1507
1507
1508
1508
1509
1509
1510
1510
1511
1511
1512
1512
1513
1513
1514
1514
1515
1515
1516
1516
1517
1517
1518
1518
1519
1519
1520
1520
1521
1521

```

```

26     case 4:
27         elementType = TYPE_VEC4UI8;
28         if(normalize) glImageFormat.desiredInternalFormat = GL_RGBA8;
29         else          glImageFormat.desiredInternalFormat = GL_RGBA8UI;
30         break;
31     default:
32         assert(0&&"should never end here");
33     }
34     break;
35     case 16:
36         glImageFormat.channelDataType = GL_UNSIGNED_SHORT;
37         if(normalize) clImageFormat.image_channel_data_type = ←
38             CL_UNORM_INT16;
39         else          clImageFormat.image_channel_data_type = CL_UNSIGNED_INT16←
40             ;
41     switch(elementInfo.numChannels)
42     {
43     case 1:
44         elementType = TYPE_UINT16;
45     /*... continuing that way for many many lines...*/

```

Listing 9: API- und Buffertyp-abhängige Operationen auf Buffern – Implementierungen durch die Buffer-Klasse

```

1 void Buffer::generateGL()
2 {
3     glGenBuffers(1, &mGraphicsBufferHandle);
4 }
5 void Buffer::generateCL()
6 {
7     mComputeBufferHandle = cl::Buffer(
8         PARA_COMP_MANAGER->getCLContext(),
9         //TODO check performance and interface "set-ability" of ←
10        CL_MEM_READ_ONLY, CL_MEM_WRITE_ONLY
11        CL_MEM_READ_WRITE,
12        mBufferInfo->bufferSizeInByte,
13        NULL,
14        PARA_COMP_MANAGER->getLastCLErrorPtr()
15    );
16 }
17 void Buffer::generateCLGL()
18 {
19     mComputeBufferHandle = cl::BufferGL(
20         PARA_COMP_MANAGER->getCLContext(),
21         //TODO check performance and interface "set-ability" of ←
22         CL_MEM_READ_ONLY, CL_MEM_WRITE_ONLY
23         CL_MEM_READ_WRITE,
24         mGraphicsBufferHandle,
25         PARA_COMP_MANAGER->getLastCLErrorPtr()
26     );
27     PARA_COMP_MANAGER->registerSharedBuffer(mComputeBufferHandle);
28 }
29 void Buffer::bindGL()
30 {
31     glBindBuffer(mGlBufferTargetEnum, mGraphicsBufferHandle);
32 }

```

```

33     glBufferData(
34         //which target?
35         mGlBufferTargetEnum,
36         // size of storage
37         mBufferInfo->bufferSizeInByte,
38         //data will be passed in setData()
39         NULL,
40         //draw static if not modded, dynamic otherwise ;
41         mContentsAreModifiedFrequently ? GL_DYNAMIC_DRAW : GL_STATIC_DRAW);
42     }
43     void Buffer::writeGL(const void* data)
44     {
45         glBufferSubData(mGlBufferTargetEnum,0,mBufferInfo->bufferSizeInByte,←
46                         data);
47     }
48     void Buffer::writeCL(const void* data)
49     {
50         PARA_COMP_MANAGER->getCommandQueue().enqueueWriteBuffer(
51             static_cast<cl::Buffer&>(mComputeBufferHandle),
52             PARA_COMP_MANAGER->getBlockAfterEnqueue(),
53             0,
54             mBufferInfo->bufferSizeInByte,
55             data,
56             0,
57             PARA_COMP_MANAGER->getLastEventPtr()
58         );
59     }
60     void Buffer::readGL(void* data)
61     {
62         glGetBufferSubData(mGlBufferTargetEnum,0,mBufferInfo->bufferSizeInByte,←
63                         data);
64     }
65     void Buffer::readCL(void* data)
66     {
67         PARA_COMP_MANAGER->getCommandQueue().enqueueReadBuffer(
68             static_cast<cl::Buffer&>(mComputeBufferHandle),
69             PARA_COMP_MANAGER->getBlockAfterEnqueue(),
70             0,
71             mBufferInfo->bufferSizeInByte,
72             data,
73             0,
74             PARA_COMP_MANAGER->getLastEventPtr()
75         );
76     }
77     void Buffer::copyGLFrom(GraphicsBufferHandle bufferToCopyContentsFrom)
78     {
79         //bind other buffer as read target
80         GUARD(glBindBuffer(mGlBufferTargetEnum, bufferToCopyContentsFrom));
81         //bind own buffer as write target;
82         GUARD(glBindBuffer(GL_COPY_WRITE_BUFFER, mGraphicsBufferHandle));
83         GUARD(glCopyBufferSubData(
84             mGlBufferTargetEnum, //operator==() asserts that buffer types are ←
85             //equal, so mGlBufferTargetEnum == rhs.mGlBufferTargetEnum
86             GL_COPY_WRITE_BUFFER,
87             0,
88             0,
89             mBufferInfo->bufferSizeInByte);
90     }
91     void Buffer::copyCLFrom(ComputeBufferHandle& bufferToCopyContentsFrom)
92     {

```

```

92     PARA_COMP_MANAGER->getCommandQueue().enqueueCopyBuffer(
93         static_cast<const cl::Buffer&>(bufferToCopyContentsFrom),
94         static_cast<const cl::Buffer&>(mComputeBufferHandle),
95         0,
96         0,
97         mBufferInfo->bufferSizeInByte,
98         0,
99         PARA_COMP_MANAGER->getLastEventPtr()
100    );
101 }
102 void Buffer::freeGL()
103 {
104     glDeleteBuffers(1, &mGraphicsBufferHandle);
105 }

```

Listing 10: API- und Buffertyp-abhängige Operationen auf Buffern – teilweise Implementationen für alle Textur-Typen durch die abstrakte Texture-Klasse

```

1 //not pure, as all the same for every texture;
2 void Texture::generateGL()
3 {
4     glGenTextures(1, &mGraphicsBufferHandle);
5 }
6 //the two non-symmetric GL-only routines:
7 //non-pure, as binding is all the same, only the enum is different, and ←
8 //that can be looked up in mTextureInfo;
9 void Texture::bindGL()
10 {
11     glBindTexture(mTextureInfoCastPtr->textureTarget, mGraphicsBufferHandle←
12         );
13 }
14 //can be non-pure, as clEnqueueWriteImage is quite generic;
15 void Texture::writeCL(const void* data)
16 {
17     cl::size_t<3> origin; origin[0]=0;origin[1]=0;origin[2]=0;
18     cl::size_t<3> region;
19     region[0] = mTextureInfoCastPtr->dimensionExtends.x;
20     region[1] = mTextureInfoCastPtr->dimensionExtends.y;
21     region[2] = mTextureInfoCastPtr->dimensionExtends.z;
22     PARA_COMP_MANAGER->getCommandQueue().enqueueWriteImage(
23         static_cast<cl::Image&>(mComputeBufferHandle),
24         PARA_COMP_MANAGER->getBlockAfterEnqueue(),
25         origin,
26         region,
27         0,
28         0,
29         //BAAAD haxx due to a bug in the ocl-c++-bindings: a otherwise ←
30         //const-param is non-const here,
31         //although it is const in the wrapped c-function 0_o
32         const_cast<void*>( data ),
33         0,
34         PARA_COMP_MANAGER->getLastEventPtr()
35     );
36 }
37 //non-pure, as glGetTexImage is quite generic :); at least one generic GL←
38 //function ;(
39 //BUT:
40 //must be overridden with an exception-throw-implementation for certain ←
41 //concrete Texture classes

```

```

37 //((namely multisample textures), which dont seem to be writable, copyable←
38 //ar readable
39 //must also be overridden by CubeMap class, as the read call must happen ←
40 //six times;
41 void Texture::readGL(void* data)
42 {
43     glGetTexImage(mTextureInfoCastPtr->textureTarget,
44                 0,
45                 mTextureInfoCastPtr->glImageFormat.channelOrder,
46                 mTextureInfoCastPtr->glImageFormat.channelDataType,
47                 data
48                 );listing
49 }
50 //can be non-pure, as clEnqueueReadImage is quite generic;
51 void Texture::readCL(void* data)
52 {
53     cl::size_t<3> origin; origin[0]=0;origin[1]=0;origin[2]=0;
54     cl::size_t<3> region;
55     region[0] = mTextureInfoCastPtr->dimensionExtends.x;
56     region[1] = mTextureInfoCastPtr->dimensionExtends.y;
57     region[2] = mTextureInfoCastPtr->dimensionExtends.z;
58
59     PARA_COMP_MANAGER->getCommandQueue().enqueueReadImage(
60         static_cast<cl::Image&>(mComputeBufferHandle),
61         PARA_COMP_MANAGER->getBlockAfterEnqueue(),
62         origin,
63         region,
64         0,
65         0,
66         data,
67         0,
68         PARA_COMP_MANAGER->getLastEventPtr()
69     );
70 }
71 //as it seems that a generic copying of many texture types can happen in ←
72 //an agnostic way
73 //via two helper FBOs, this can be non-pure;
74 //BUT:
75 //must be overridden with an exception-throw-implementation for certain ←
76 //concrete Texture classes
77 //((namely multisample textures), which dont seem to be writable, copyable←
78 //or readable;
79 //must also be overridden by CubeMap, as there are six color attachments ←
80 //to copy; or should one work with
81 //glFrameBufferTextureFace ? well see in the future ;
82 void Texture::copyGLFrom(GraphicsBufferHandle bufferToCopyContentsFrom)
83 {
84     //TODO implement when FBO class exists;
85     throw(BufferException("Texture ::copyGLFrom: sorry, for copying, FBOs ←
86     //are needed, and the still must be implemented "));
87 }
88 //can be non-pure, as clEnqueueCopyImage is quite generic;
89 void Texture::copyCLFrom( ComputeBufferHandle & bufferToCopyContentsFrom)
90 {
91     cl::size_t<3> origin; origin[0]=0;origin[1]=0;origin[2]=0;
92
93     cl::size_t<3> region;
94     region[0] = mTextureInfoCastPtr->dimensionExtends.x;
95     region[1] = mTextureInfoCastPtr->dimensionExtends.y;
96     region[2] = mTextureInfoCastPtr->dimensionExtends.z;
97
98     PARA_COMP_MANAGER->getCommandQueue().enqueueCopyImage(

```

```

92     //source
93     static_cast<const cl::Image&>(bufferToCopyContentsFrom),
94     //destination
95     static_cast<const cl::Image&>(mComputeBufferHandle),
96     origin,
97     origin,
98     region,
99     0,
100    PARA_COMP_MANAGER->getLastEventPtr()
101   );
102 }
103 //can be non-pure, as glDeleteTextures() applies to every texture type :)
104 void Texture::freeGL()
105 {
106   glDeleteTextures(1, &mGraphicsBufferHandle);
107 }
```

Listing 11: API- und Buffertyp-abhängige Operationen auf Buffern – restliche Implementationen durch die Texture2D-Klasse

```

1 void Texture2D::generateCLGL()
2 {
3   mComputeBufferHandle = cl::Image2DGL(
4     PARA_COMP_MANAGER->getCLContext(),
5     CL_MEM_READ_WRITE,
6     mTextureInfoCastPtr->textureTarget,
7     0,
8     mGraphicsBufferHandle,
9     PARA_COMP_MANAGER->getLastCLErrorPtr()
10    );
11 }
12 void Texture2D::allocGL()
13 {
14   glTexImage2D(
15     mTextureInfoCastPtr->textureTarget,
16     0,
17     mTextureInfoCastPtr->glImageFormat.desiredInternalFormat,
18     mTextureInfoCastPtr->dimensionExtends.x,
19     mTextureInfoCastPtr->dimensionExtends.y,
20     0,
21     mTextureInfoCastPtr->glImageFormat.channelOrder,
22     mTextureInfoCastPtr->glImageFormat.channelDataType,
23     //dont set data yet, just alloc mem
24     0
25   );
26 }
27 }
28 void Texture2D::writeGL(const void* data)
29 {
30   glTexSubImage2D(
31     mTextureInfoCastPtr->textureTarget,
32     0,
33     0,0,
34     mTextureInfoCastPtr->dimensionExtends.x,
35     mTextureInfoCastPtr->dimensionExtends.y,
36     mTextureInfoCastPtr->glImageFormat.channelOrder,
37     mTextureInfoCastPtr->glImageFormat.channelDataType,
38     data
39   );
40   if(mTextureInfoCastPtr->isMipMapped)
```

```
41 {  
42     glGenerateMipmap( mTextureInfoCastPtr->textureTarget );  
43 }  
44 }
```

Literatur

- [APS00] Michael Ashikhmin, Simon Premoze, and Peter Shirley. A microfacet-based brdf generator, 2000.
- [BMF07] Robert Bridson and Matthias Müller-Fischer. Fluid simulation: Siggraph 2007 course notes. In *ACM SIGGRAPH 2007 courses*, SIGGRAPH '07, pages 1–81, 2007.
- [Ebe04] David H. Eberly. *Game Physics*. Interactive 3D Technology. Morgan Kaufman, 2004.
- [GSSP10] Prashant Goswami, Philipp Schlegel, Barbara Solenthaler, and Renato Pajarola. Interactive SPH simulation and rendering on the GPU. In *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '10, pages 55–64. Eurographics Association, 2010.
- [HSO07] Mark Harris, Shubhabrata Sengupta, and John D. Owens. Parallel Prefix Sum (Scan) with CUDA. In Hubert Nguyen, editor, *GPU Gems 3*, chapter 39. Addison-Wesley Professional, 2007.
- [Khr09] Khronos Group. *The OpenCL Specification*, version: 1.0, document revision: 48 - 10/6/09 edition, 2009.
- [Khr10] Khronos Group. *The OpenGL Graphics System: A Specification*, version 4.1 (core profile) - july 25, 2010 edition, 2010.
- [LG08] Scott Le Grand. Broad-Phase Collision Detection with CUDA. In Hubert Nguyen, editor, *GPU Gems 3*, chapter 32. Addison-Wesley Professional, 2008.
- [MCG03] Matthias Müller, David Charypar, and Markus Gross. Particle-Based Fluid Simulation for Interactive Applications. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*, SCA '03, pages 154–159. Eurographics Association, 2003.
- [Pes09] Franz Peschel. Simulation und Visualisierung von Fluiden in einer Echtzeitanwendung mit Hilfe der GPU. Studienarbeit, Universität Koblenz-Landau, apr 2009.
- [Sta99] Jos Stam. Stable fluids. pages 121–128, 1999.
- [Sta03] Jos Stam. Real-time fluid dynamics for games, 2003.
- [Ste07] Thomas Steil. Efficient Methods for Computational Fluid Dynamics and Interactions. Diplomarbeit, Universität Koblenz-Landau, dec 2007.

- [vdLGS07] Wladimir J. van der Laan, Simon Green, and Miguel Sainz. Screen Space Fluid Rendering with Curvature Flow. In Eric Haines, Morgan McGuire, Daniel G. Aliaga, Manuel M. Oliveira, and Stephen N. Spencer, editors, *SI3D*, pages 91–98. ACM, 20092007.