

Department of Physics and Astronomy
University of Heidelberg

Master thesis
in Computer Engineering
submitted by
Dennis Sebastian Rieber
born in Balingen
2017

Characterization Of GPU Communication

A Scientific Exploration

This Master thesis has been carried out by Dennis Sebastian Rieber
at the
ZITI, Computer Engineering Group
under the supervision of
JProf. Dr. Holger Fröning

Abstract

In HPC world, the communication of massive parallel workloads in distributed memory systems thoroughly analysed. GPU hardware and software continues to scale in parallelism, yet the intra-application communication hasn't been a topic of research. However, the understanding of this intra-application communication is of fundamental importance for multi-GPU scaling of applications.

This work uses compiler extensions to instrument GPU applications for tracing. The instrumentation adds tracing for all load and store instructions targeting global memory in a GPU. Data is dynamically transferred from the device to host at run time, through a producer-consumer controlled buffer.

Using the generated traces, communication across BSP supersteps is analysed. The analysis shows messages between supersteps are dominated by few transfer sizes. The number of communication partners and memory strides in communication depend on the application. Irregular applications display irregularities in the amount of data and number of communication partners across several supersteps, while regular applications do not. All applications show a tendency towards collective communication.

In der Welt des High-Performance Computing ist die Kommunikation von distributed-memory Anwendungen ausgiebig erforscht. Die Parallelität von GPU Hardware und Software wächst immer weiter, aber die Kommunikation innerhalb von Anwendungen wurde bis jetzt noch nicht untersucht. Das Verständnis dieser Kommunikation ist aber essenziell für Multi-GPU Skalierung von Anwendungen.

Diese Arbeit nutzt Compiler-Erweiterungen, um GPU Anwendungen so zu instrumentieren, dass diese Traces erzeugen. Das Tracing geschieht für alle Speicherzugriffe auf den Global Memory einer GPU. Die Daten werden dynamisch von der GPU auf den Host übertragen, über einen Erzeuger-Verbraucher gesteuerten Speicher.

Die generierten Daten werden dazu verwendet, die Kommunikation über BSP Supersteps zu analysieren. Eine Analyse zeigt die Übertragungen zwischen Kernen sind von einigen wenigen Größen dominiert. Die Anzahl der Kommunikationspartner und Größe des Memory Strides während der Kommunikation hängen von der Anwendung ab. Irreguläre Anwendungen zeigen Unregelmäßigkeiten im übertragenen Datenvolumen und der Anzahl der Kommunikationspartner über mehrere Supersteps hinweg, wohingegen reguläre Anwendungen das nicht zu tun. All Anwendungen zeigen eine Tendenz zur kollektiven Kommunikation.

Contents

1 Motivation	1
1.1 Goals	2
1.2 Outline	2
2 Background	4
2.1 Bulk Synchronous Parallel	4
2.2 General Purpose GPU Computing	5
2.2.1 Execution Model	5
2.2.2 Memory Model	5
2.3 LLVM	7
2.3.1 LLVM IR Program Representation	7
2.3.2 Static Single Assignment	8
2.3.3 LLVM Pass	10
2.4 Compiler Techniques	11
2.4.1 Data Flow Analysis	11
2.4.2 Function Specialization	12
3 Related Work	13
3.1 GPU Application Characterization	13
3.2 GPU Code Instrumentation	14
4 Methodology	16
4.1 GPU Shared Memory Communication	16
4.2 Analysis Space Exploration	18
4.2.1 Analysis Methods	18
4.2.2 Caveats	18

5 Dynamic Instrumentation Using Code Analysis And Transformation For Detailed GPU Communication Analysis	20
5.1 Trace Setup	20
5.1.1 Build Process	21
5.1.2 Trace Format	22
5.2 Source to Source Compilation in Clang	22
5.3 Code Transformation in LLVM	23
5.3.1 Address Space Analysis	25
5.3.2 Instrumentation	29
5.4 Tracing Process	30
5.4.1 On-Device Producer	31
5.4.2 Host Consumer	34
6 Evaluation and Analysis of GPU Communication	35
6.1 Evaluated Applications	35
6.2 Tier I Analysis	36
6.2.1 Total Data Volume and Communication Fraction	36
6.2.2 Transfer Size CDF	37
6.2.3 CTA In/Out-Degree	37
6.2.4 Bisection Volume	38
6.2.5 Communication Stride	39
6.2.6 Density Map	40
6.3 Tier II Analysis	42
6.3.1 Kernel communication Evolution	42
6.3.2 Superstep Distance	42
6.4 Tier III Analysis	43
6.4.1 CTA connectivity evolution	43
7 Conclusion	45
7.1 Future Work	46
A Instrumentation	IV
Appendices	IV
B Post Processing	VI

List of Figures

2.1	GPU thread and memory hierarchy. Image from []	6
2.2	non-SSA vs. SSA code	8
2.3	Branching program and resulting CFG. Not in SSA form.	9
2.4	Branching program and resulting CFG. Now in SSA form, with ϕ -function resolving merging CFG paths	9
2.5	Def-Use connections of variable a1, displayed as dotted lines.	10
3.1	Lynx Framework	15
4.1	CTA Communication	17
5.1	Structure of an instrumented application. Segments added during instrumentation are highlighted in blue.	21
5.2	Trace record format	22
5.3	Trace file format	22
5.4	Example for address space lowering in LLVM IR	24
5.5	ϕ -node Limitations	28
5.6	Example for ambiguous function address spaces	29
5.7	Producer CFG	32
5.8	Device Producer, naive approach	32
5.9	Device Producer, warp scope	33
6.1	Total data volume and ratio of communication to total of all writes.	36
6.2	Message Size CDF Plots	37

LIST OF FIGURES

6.3	Number of incoming and outgoing communication partners of each CTA	38
6.4	Total and relative bisection Volume, separated by dimensions	39
6.5	Strides of loads and stores used in communication, by CTA	40
6.6	Transfer Density Plots	41
6.7	Volume of communication loads and stores in each iteration	43
6.8	Volume of communication for <i>bfs</i> application, separated by superstep distance.	44
6.9	Mean out-degree of CTAs, per iteration	44

List of Tables

2.1 GPU Memory [?, 2.3]	6
5.1 Memory operation descriptor table	31
6.1 Application used for communication evaluation	35

Chapter 1

Motivation

The communication of HPC applications in distributed memory systems with MPI is thoroughly researched in many publications, like ... or With respect to the BSP modell, GPU computing has many similarities to computing in distributed memory systems. Each CTA computes independently and information is stored in memory. CTAs in subsequent kernels read this data, creating patterns of communication. This resembles MPI, as each node computes locally, and then communicates and synchronizes with other nodes, before the next part of computation. Yet, intra-application communication in GPUs has not been researched, despite the fact that General Purpose GPU (GPGPU) computing is a well established part of high performance computing, as many Top500 systems [?] use GPUs to accelerate computation.

In GPUs, parallel hardware continues scaling with every generation, from 512 CUDA cores in Fermi [?] to 3840 CUDA cores in the current Pascal generation [?]. This scaling is supported by CUDA's programming model, allowing performance portable applications. With increasing potential parallelism and concurrency, communication patterns, volumes and complexity also increase or change. An understanding of intra-application communication is fundamental for the scaling of applications to multiple GPUs.

To create this understanding, an in-depth knowledge about the communication is required. The knowledge can be generated using different techniques. Spafford et al. [?] categorized these techniques and presented several metrics for evaluation, including speed and accuracy. For high accuracy requirements, GPGPU simulators provide cycle-accurate simulation results for an application. However, it takes very long to simulate every hardware cycle in software which makes it a time-consuming process. Another issue with simulation is the implementation of different hardware generation. GPGPU-Sim, for example, only supports Fermi architectures. [?]

Analytical modelling uses mathematical models, trying to create an approximation on the ap-

plications behaviour. While this can be done quickly, accuracy is usually lacking, especially for dynamic or data-depended applications.

MPI communication is characterized using traces. An application instrumented for tracing logs information on a desired metric and stores the data for later analysis. Instrumented code is faster than simulations as it runs on the actual hardware and can be used on any hardware generation. Although slower than analytical modelling, it is more accurate since it is generated while the actual application executes, using a real dataset. These features make instrumentation and tracing the tool of choice for this thesis.

1.1 Goals

This thesis has two main goals. First, develop tooling for instrumentation of GPU applications to perform dynamic global memory tracing. The instrumentation will happen at compile time using custom plugins for Clang and LLVM to transform the source code of the original application. With this setup, traces can be generated for any application providing source code, across different generations of Nvidia hardware. A loss in performance resulting from the traces will be tolerated.

The second goal is to use the generated traces for CTAs and kernels communicate analysis. The communication will be analysed on a qualitative level by classifying how the data is exchanged and on a quantitative level with metrics like volume, density and distance to characterize kernel and CTA interactions.

1.2 Outline

Chapter 2 presents technological background information. It will introduce BSP, GPUs, LLVM, SSA and the compiler techniques used to transform the code of the original application.

Chapter 3 discusses work related to this Thesis. It will explore work in the field of code instrumentation and GPU performance analysis.

Chapter 4 defines communication in the context of this work and explores the analysis space.

Chapter 5 presents how applications are instrumented using AST manipulation and deeper code analysis in LLVM's intermediate representation (IR).

Chapter 6 presents the set of applications that will be analysed, the different metrics are explained in detail and the results of the analysis are discussed.

Chapter 7 concludes and summarizes the results of this thesis and discusses possible future directions this project could take.

Chapter 2

Background

2.1 Bulk Synchronous Parallel

Bulk Synchronous Parallel (BSP) is a bridging model introduced by L.G. Valiant [?]. A bridging model describes a conceptual understanding of hardware for the purpose software development. The most basic of these models is the von Neumann architecture, which is still used as the basic understanding of sequential hardware in modern CPUs. In the wake of parallel computing, BSP was proposed to fulfil the same roll for the understanding of parallel hardware. The model does not specify how any of its parts should be realised in hardware or software, since it's purpose is to provide a common ground for hardware and software developers. BSP fundamental principal is the separation of computation and communication. A computer modelled after BSP requires

1. a set of components or execution units, performing the computational work of the algorithm.
2. a router to convey information between the components.
3. a facility to synchronize some or all of the components in steady intervals.

Each component performs local computations and then shares information with the other components via the router. The synchronization guarantees that all components completed computation and communication. The sequence of computation, communication and synchronization is called a superstep. One or more supersteps are used to implement an algorithm. The model allows excluding components from synchronization, if the algorithm allows or requires it.

Although all computations are supposed to be local, BSP allows concurrent read and concurrent write access to a shared memory between the components. But only if the underlying memory system guarantees coherent and consistent resolution of access conflicts.

2.2 General Purpose GPU Computing

GPGPU computing utilizes GPUs for computations other than graphics. With CUDA, NVidia offers a C-based, proprietary language to write application for NVidia GPUs. GPUs are so called "many-core" processors, with thousands of compute units (CUDA cores). CUDA cores are clustered in Streaming Multiprocessors(SM). Code executed on a GPU is capsule in a compute kernel. Figure 2.1 is an overview of the features detailed in the next sections. Nvidia also offers an assembler-like language called PTX, that is not architecture generation specific. [?, 4.1-4.2]

2.2.1 Execution Model

Threads in CUDA are hierarchically organized. Groups of up to 1024 threads are organized in a 3D block, called "Collaborative Thread Array" (CTA). The execution of one compute kernel consists of many CTAs, organized in a 3D grid. During a kernel execution the CTAs are scheduled to a SM in a non-deterministic fashion, based on available resources. Once a CTA is scheduled, it can not be pre-empted and executes until it is finished. [?, 2.1 - 2.2]

The threads inside a CTA are organized in groups of 32, called warp which execute in a Single Instruction Multiple Threads (SIMT) fashion. If the one or more threads in a warp have a branch in the Control-Flow-Graph (CFG), all threads in the warp execute the branch, but only the ones supposed to work create side-effects. The shared instruction stream continues if all threads finished a branch. [?, 4.1]

A kernel usually consists of more CTAs than a GPU can execute concurrently, to utilize parallel slackness and create performance portable applications. Non-deterministic scheduling on the GPU prevents predictions on the CTA execution order, making CTA interaction in a kernel a deadlock hazard. Therefore interactions of threads belonging to different CTAs are not allowed. The execution model allows threads inside the same CTA to interact using Shared Memory, which is located in the SM and private for each CTA. [?, 5]

Consecutive kernels are executed sequentially and in calling order. To improve concurrent utilizations of all hardware units, CUDA offers streams. Each stream has a sequential execution order, but multiple streams can be executed concurrently. [?, 5]

2.2.2 Memory Model

GPUs have a hierarchical memory model, with different address spaces. Table 2.1 lists all available memories and their properties. For this work we will focus on global memory, as it is

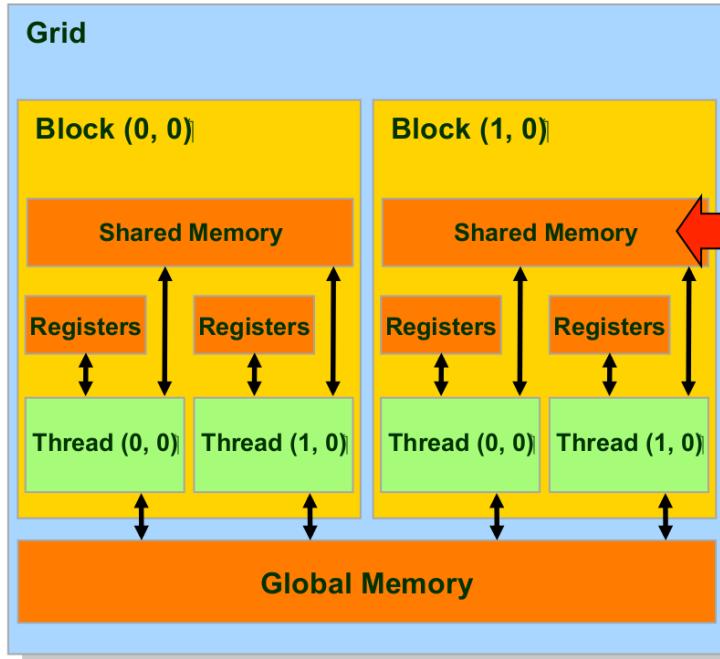


Figure 2.1: GPU thread and memory hierarchy. Image from []

Memory	Scope	Access	Bandwidth	Latency	Capacity	Location
Register	Thread	RAM	high	low	32bit	SM
Shared	CTA	RAM	high	low	< 48k	SM
Global	Kernel	RAM	high	high	< 20GB	device
Texture/Constant	Kernel	ROM	high	high	< 20GB	device
Host-Mapped	Kernel+CPU	RAM	low	high	> 20GB	host

Table 2.1: GPU Memory [?, 2.3]

the only memory allowing modification by a kernel, which persist beyond the kernel completion boundary. Host-Mapped memory falls in the same category as global, and for our purposes is treated as global memory. Global address space is a shared memory for all kernels running a GPU.

It is important to note, that CUDA only guarantees a consistent state of global memory, after the kernel completion boundary. There are no guarantees on global memory consistency during the execution of a kernel, which is another reason CTA interaction at kernel run-time is not intended. [?, B5-B7]

2.3 LLVM

LLVM is a compiler project providing tools for full-stack compiler development. While originally written for C/C++, nowadays front-ends for many languages are provided and back-ends for many different architectures are available. The C/C++ compiler using the LLVM tool-chain is called Clang. Since 2015 Clang and the LLVM tool-chain support CUDA code and can produce PTX files for GPU execution [?]. LLVM uses an intermediate representation (IR) for code analysis and optimization (more on this in 2.3.1). Code compiled with Clang is translated into IR and then further processed. [?]

For this work, we will distinguish between three major components in the tool-chain:

- The **front-end** takes care of pre-processing, lexical, and syntactical analysis of the original code. Clang uses an Abstract Syntax Tree (AST) to represent the original code. This AST is accessible via an API and additional plug-ins can be added to the tool-chain. After the lexical and syntactical analysis are complete, the code is translated into IR.
- The **optimizer** uses the IR to analyse and then transform the code. The IR is accessible with an API and allows the introduction of additional modules for analysis or transformation. These modules are called a "pass" (details in section 2.3.3).
- The **back-end** includes the linker and architecture depended code generation. This part of the toolchain is not relevant for this work and is only mentioned for completeness.

This overview gives an sufficient understanding of the LLVM stack for this work and in the following sections the concepts of IR and passes are explored in more depth.

2.3.1 LLVM IR Program Representation

LLVM's IR is a typed RISC instructions set with a load/store memory architecture. The IR is architecture agnostic and provides an infinite set of virtual, typed registers. The available primitive language-independent types are: (un)signed integer (8-64 bit), single and double precision floating point, and Boolean. Derived from these primitive types are the aggregate types pointers, arrays, structures, and functions. In LLVM IR it is possible to transform form any type into an arbitrary other type, using the `cast` instruction. [?]

Any kind of pointer arithmetic and access to aggregate types is performed by the `getelementptr` (`gep`) instruction, which preserves type information, is machine and language independent, and returns a pointer. Using `gep` allows the IR to keep load/store instructions clean and only use the direct address to access element. [?]

```
1 a = 5
2 b = a + 1
3 a = 2
4 c = a + 1
```

```
1 a1 = 5
2 b1 = a1 + 1
3 a2 = 2
4 c1 = a2 + 1
```

Figure 2.2: The left hand side is non-SSA program text, on the right hand side is the corresponding SSA representation. Variable 'a'’s value is changed in line three. In SSA, a new definition for 'a' is created . It is customary to use a counting index for successive redefinition of variables in SSA.

One important part of LLVM’s code transformations are canonicalization and lowering, which happenin the front-end. One example of lowering is the `gep` instruction, that brings any kind of language intrinsic for address access into the same form. Canonicalization happens for logical constructs in the program’s CFG. For example, any loop in a program is brought into the same canonical form before any analysis on the loop is performed. The reason for this is the simplification and optimization of the passes analysing and transforming the program. [?]

2.3.2 Static Single Assignment

LLVM’s IR is represented in Static Single Assignment (SSA) form. Informally, it can be defined as "A program is defined to be in SSA form if each variable is a target of exactly one assignment statement in the program text" [?, p. 6].

ϕ Functions

A simple example of a program and the corresponding SSA form is displayed in figure 2.2. This example has no branches and would result in a strictly sequential CFG. Figure 2.3 shows how a branching program creates divergence in the CFG. This creates additional basic blocks, which then merge back together at the end of the branch. The program used in 2.3 is not in SSA form, as variable `b` is defined at two different places in the program text. One of the most integral concepts of SSA is used to resolve situations like this. The ϕ -function (sometimes called ϕ -node), is a "pseudo assignment function" [?] which resolves multiple incoming values at the location a CFG merges back together by defining a new variable that is used in the later program flow. This is displayed in figure 2.4. [?]

The ϕ -function has n parameters for the n incoming edges to the basic block. For the definition of the new variable, it selects the value belonging to edge active at runtime. It is important to keep in mind that ϕ -node are a helper, that is not actually present in the final program. After analysis and transformation are finished by the optimizer, the SSA form is de-constructed by the

```

1 if (a = 5)
2     b = 3
3 else
4     b = 5
5 print(b)
6

```

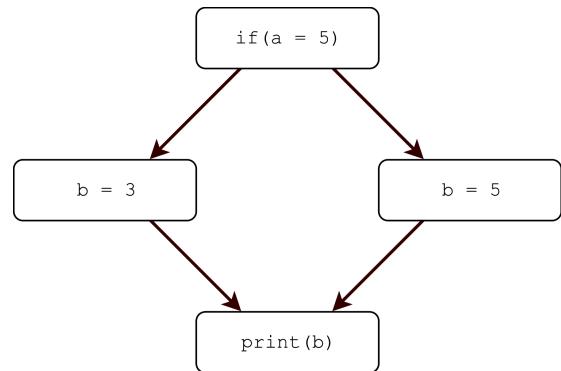


Figure 2.3: Branching program and resulting CFG. Not in SSA form.

```

1 if (a1 = 5)
2     b1 = 3
3 else
4     b2 = 5
5 b3 = phi(b1,b2)
6 print(b3)
7

```

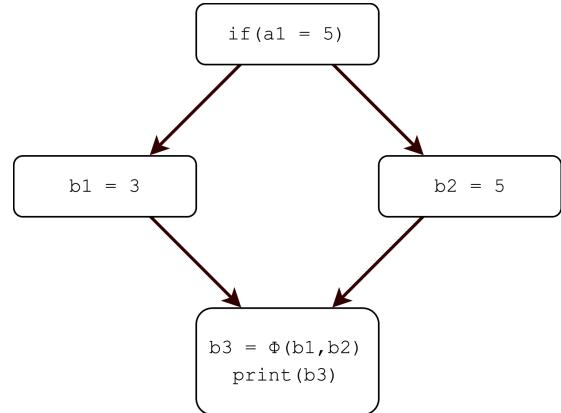


Figure 2.4: Branching program and resulting CFG. Now in SSA form, with ϕ -function resolving merging CFG paths

back-end. Now, as values are mapped to registers and memory locations that can be overwritten, the ϕ -nodes and SSA form are no longer necessary. [?]

Def-Use and Use-Def Chains

An important data structure in compiler analysis are def-use and use-def chains. The former is a list of every use a definition has. Use-def chains point backwards to all definitions of variable, from the perspective of a use. [?]

Due to the single definition variables of SSA, use-def chains are a single name, or element. Def-use chains can be built easily because they can be constructed through the single element use-def chains. We look at each use in the program and add the user to the definition's def-use list. Def-Use connections enable fast forward travelling through a CFG, because one is not bothered with code that is not part of the def-use chain. Figure 2.5 shows an example of def-use

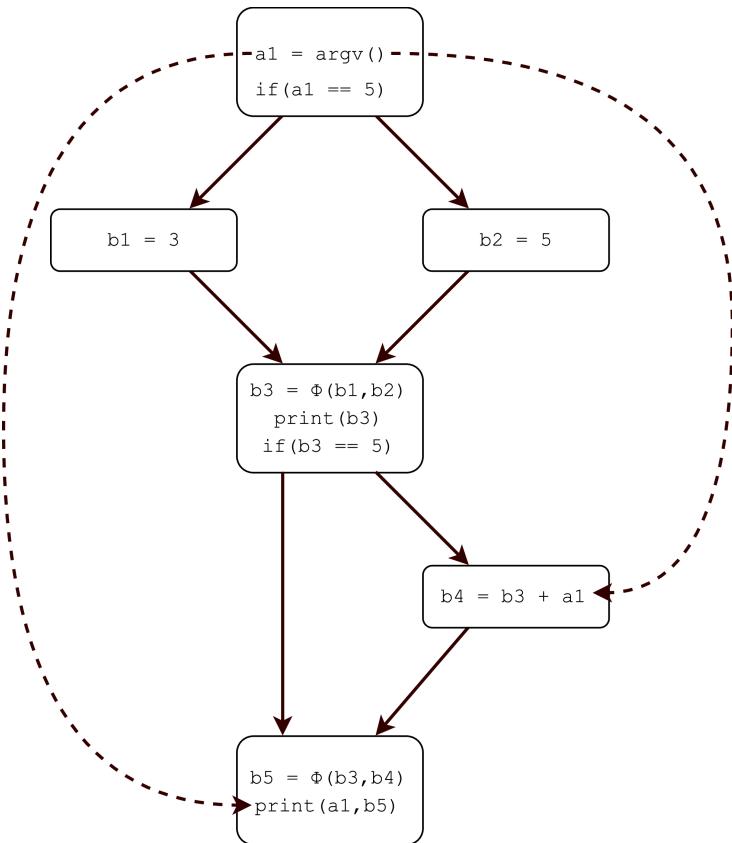


Figure 2.5: Def-Use connections of variable `a1`, displayed as dotted lines.

connections for the variable `a1`. The use-def connections would be the inversion of the dotted arrows. [?]

2.3.3 LLVM Pass

A pass in LLVM is a piece of software responsible for an analysis or transformation of IR. Passes are distinguished into said categories. An analysis pass creates or computes information on the piece of IR it is running on, which can be used later by other passes. It doesn't modify the IR. As one pass may depend on the result of another analysis, passes are scheduled accordingly by the pass manager. A transformation pass alters the IR in place and does not produce any results. A transformation might trigger the re-run of an analysis as changes to the IR might invalidate the original results. Opposed to an analysis, transformations are not allowed to rely on results of other transformations. [?]

Scopes

A pass can run on various scopes of IR. The pass can not access IR outside this scope and in the same pass, information can not be conveyed from one scope's context to another. For example: If a pass analyses functions, each function in the IR is analysed individually, with no knowledge about other functions. [?]

- **Module** scope contains the whole compilation unit with all it's functions and definitions.
It is largest scope available.
- **Function** scope looks at every function individually. As the pass is limited to a function definition, it has no calling-context information.
- **Loop** scope only provides access to the loop head and body, including loops nested into the loop.
- **BasicBlock** can only access a single basic-block. No CFG manipulation possible.

2.4 Compiler Techniques

This section will briefly introduce terminology for two types of analysis and transformation a compiler can perform, which are relevant for this work.

2.4.1 Data Flow Analysis

Data flow Analysis is a broad field, with some shared terminology that is described here. Generally, such an analysis traverses the code, search for specific facts and optimizing the code according to the analysis results. Examples are constant propagation or zero value propagation. It is also closely related to pointer analysis, in terms of terminology and overall structure.

- **Property Space** contains the facts, the analysis can generate. Represented as partially ordered sets with anti-symmetric, transitive and reflexive relation. [?]
- A **Transfer function** classifies each visited analysis object, and determines to which set of the property space an object belongs to. [?]
- **Program Representation** is how the program is represented during the analysis. Usually the CFG is used, but more sparse representation like def-use connections can be more convenient. [?]

- **Context Sensitive** describes the attribute, that the analysis of a function is sensitive to the context the function is called in. For example, the analysis takes the functions arguments into regard. [?]
- **Flow Sensitivity** defines that the order of instructions is of importance for the analysis. [?]

2.4.2 Function Specialization

Specialization (sometime procedure cloning) create distinct versions of a functions for a specific purpose at compile time. The technique was first introduced in [?]. It resembles templates in C++, creating a function for every template used argument. A specialized function relies on facts gathered during the analysis create a specialized version. An example for this are functions with differently unrolled loops, depending on the calling context of the function.

It differs from inlining, because it doesn't interfere with the original code structure. The difference to traditional data-flow analysis based optimizations is that it's possible to create multiple specialized versions for multiple data-flow facts, that vary depending on the calling context.

Chapter 3

Related Work

3.1 GPU Application Characterization

Since GPGPU computing exists, there are publications characterizing GPU application. Generally, the analysis aspects can be separated into high-level and low-level. The former treats an application as a black-box and measures behaviour of the complete application. The latter inspects hardware level behaviour without being concerned, which part of the application is executed. Most research papers report how changes in software or hardware influence certain aspects of low-level or high-level behaviour.

High-level research often reports speed-up resulting from a change in hardware or software. Abe et al. [?] report on power efficiency and performance. High-level is also concerned with configuration, such as launch configurations and number of kernels ([?], [?]). Wang et al. [?] report on how dynamic programming affects kernel count and launch configurations of irregular workloads.

Low-level can be divided into memory and execution. Research concerned with GPU memory behaviour report L2 cache hits/efficiency ([?]), global memory load/store counters ([?], [?]) or use these numbers together with instruction replay and coalescing to create metrics like memory intensity ([?]) or regularity ([?]). Execution research is concerned with control-flow, instruction counts, and occupancy ([?], [?]).

To the best knowledge of the author, there are no publication researching intra-application interactions of kernels or CTAs.

3.2 GPU Code Instrumentation

Instrumenting existing application with addition code for memory tracing is essential for this work. A framework for code instrumentation is presented in "Lynx: A Dynamic Instrumentation System for Data-Parallel Applications on GPGPU Architectures" [?] by Farooqui et al. Lynx is a framework including a C-to-PTX JIT, that extends CUDA code on PTX level with user defined instrumentation, and a runtime, displayed in 3.1. The user writes instrumentation specification in a C-style language. The specification is translated to PTX and inserted into the original application.

The application can be instrumented on kernel, basic block and instruction level. An API offers access to CUDA constructs like Thread and CTA IDs, CTA barriers and instruction counts. To increase efficiency, it is possible that only one thread in a warp performs the specified trace operation. It is possible to access shared and global memory space. The runtime manages all buffers introduced by the instrumentation, including the movement from and to the device.

There are two reasons, Lynx is not suitable for this project. First, while it is possible to instrument global memory operations, we found no way to access details about the instrumented instruction, like target address or type size. However, address information is crucial for this work. The second reason is the runtime handling the data buffers. As the generated trace data can be very big, device memory would not suffice to hold all the data, and creating a dynamic producer-consumer buffer using Lynx is not possible. To the knowledge of the author, there is no performance analysis beyond metric counters available for GPUs.

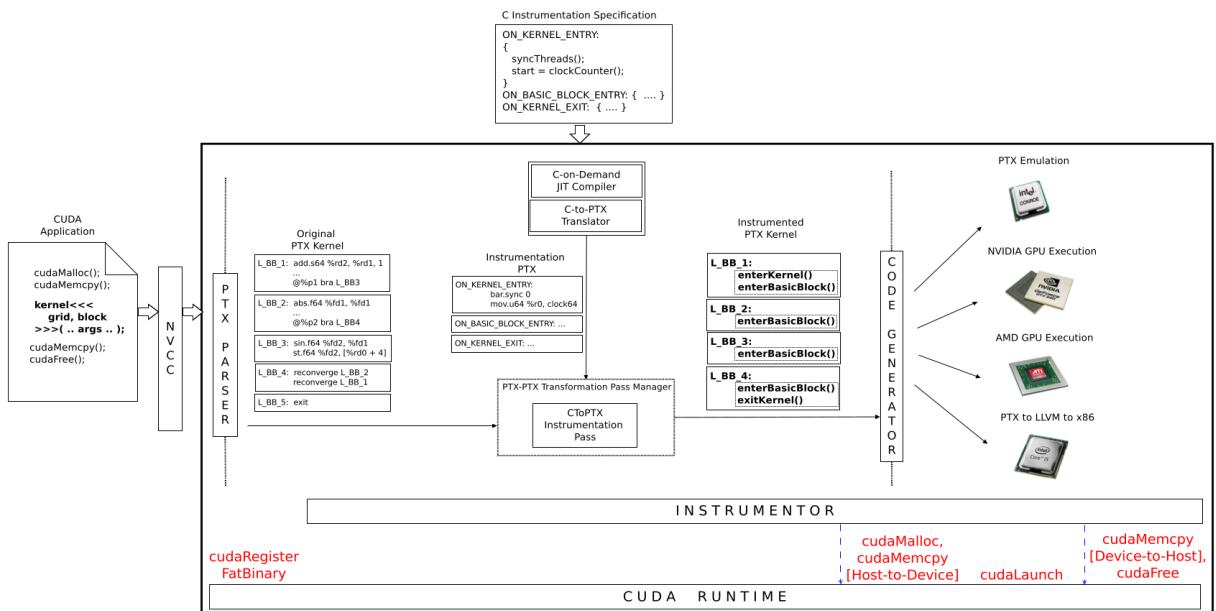


Figure 3.1: Lynx framework with all components. The original application is instrumented on PTX level, with code generated by a C-to-PTX JIT from instrumentation specification. Generated data is moved and form the device by the runtime. Image from [?]

Chapter 4

Methodology

This chapter explores how GPUs convey information using a shared memory and defines what communication means in the context of this work. Using this definition, the analysis space is described and explored.

4.1 GPU Shared Memory Communication

Other than the explicit communication routines of MPI like send and receive, GPUs convey information implicitly via the global memory on the device. The CUDA execution model and the given guarantees can be mapped to the Bulk-Synchronous-Parallel (BSP) bridging model, explained in 2.1.

- **Computation** is a kernel executed on the GPU. The local components in the model are represented by CTAs on a GPU. Although BSP allows shared memory interactions, CUDA does not, therefore each CTA is processing its own workload without the ability to directly interact with other CTAs.
- **Communication** in this step data generated in the computation step is placed in a location where it can be accessed by a component using the data in a following superstep. For GPUs this location is global memory, because it is modifiable by a kernel during execution and its modifications persist beyond kernel completion boundaries. In other words, global memory is a memory area allowing visible side-effects of a kernel execution.
- **Synchronization** on the GPU happens implicitly by the kernel completion boundary. It guarantees visibility of all global memory operations by the kernel. This implies that follow-up supersteps can see the memory modifications. Just as the barrier in BSP makes

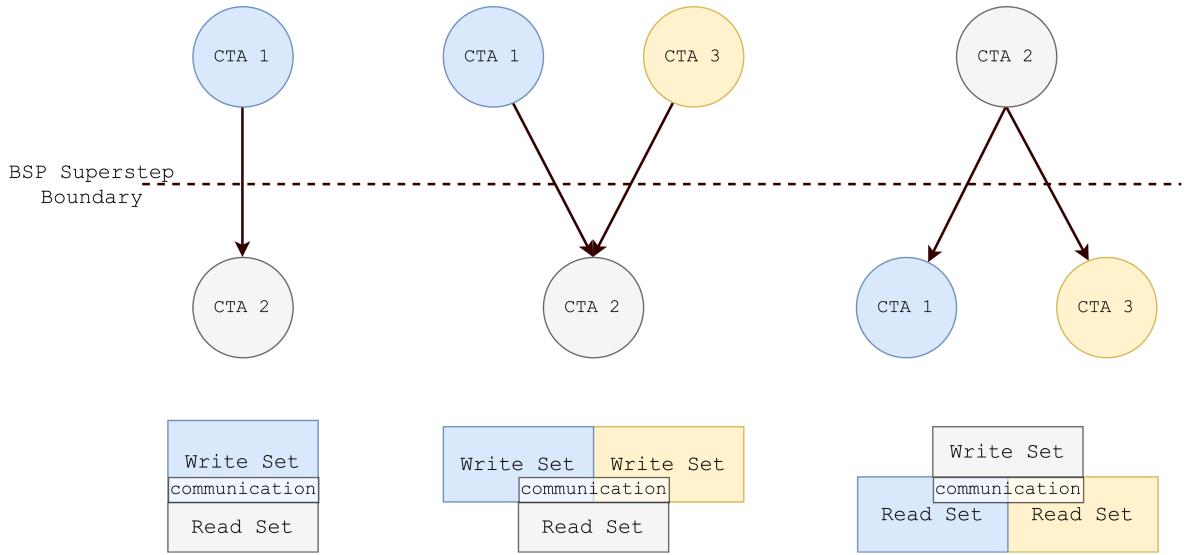


Figure 4.1: Communication of CTAs across a BSP superstep boundary, and how this corresponds to read and write sets of sinks and sources. Edges are logical representations of communication, which is actually the overlap of write and read addresses in different BSP supersteps. Different colors represent different sink/source entities.

sure the communication step has been completed by all components. Partial synchronization is only possible at stream granularity.

Not all of the side-effects mentioned above however, can be classified as communication. In the context of this work, communication is data that is written to global memory by one thread and read by another. The CUDA consistency model allows reliable communication only across kernel completion boundaries. We only consider elements that are communicated reliably.

To exemplify the definition above, figure 4.1 shows how communication between CTAs across a BSP superstep boundary translates to overlaps of write- and read-sets. The nodes above the superstep boundary are data sources, the nodes below sinks. The accumulated edge-weight is the size of the overlapping area in the write- and read-sets. The write-set is only in a later superstep reliably readable. The left-most graph is a simple point-to-point connection, the middle one resembles a gather collective and the right-most graph can be a scatter or multicast, depending on the overlap. More on collectives in 4.2.2. This representation works for all levels of granularity in the hierarchy of the GPU execution model, from a single thread to an application using multiple kernels and streams.

4.2 Analysis Space Exploration

The analysis space for this work can be separated into spatial and temporal dimensions. The spatial dimension includes the layered hierarchy of threads, CTAs and Kernels. Each hierarchy layer is the superset of the underlying layers. For example, the total data volume written by a kernel is the sum of all its CTA's writes. Granularity determines how detailed the spatial dimension is analysed. The coarsest granularity would be a complete application with multiple kernels and the finest a single thread. For this work however, CTAs are used as the finest granularity. The spatial dimension can help characterize an application, but by itself cannot describe communication.

The temporal dimension describes the series of executed BSP supersteps, which makes the time discrete. Without this dimension, there is no communication to analyse, because by our definition, communication requires at least two subsequent supersteps. Again, different granularity can be applied to the analysis. Granularity refines, as the term to describe the analysis in this dimension becomes more and more specific. The coarsest granularity ignores individual kernels and exclusively looks at the interaction between the existing supersteps. At finer granularities, the number of supersteps in the analysis may change, because only the interaction of two distinct kernels k_1 and k_2 across time t is of interest.

4.2.1 Analysis Methods

Based on the dimensions and granularities, different tiers for the analysis can be defined.

- I Kernel behaviour across all supersteps
- II Kernel interactions across two supersteps
- III CTA interactions across two supersteps

The analysis tries to avoid averaging and instead use aggregated values and relational metrics for comparison and analysis.

4.2.2 Caveats

- This work uses CTAs as the fundamental entities of communication source and sink, not threads.

- While MPI offers collective communication with clear definitions such as scatter, gather or multicast, the implicit memory communication of GPUs often can not be categorized as clearly. However the in- and out-degree of a communication entity might indicate tendencies towards a certain kind of collective. For the sake of the analysis, each collective communication ($\{n:1\}$ $\{1:n\}$ or $\{n:n\}$) can be viewed as a set of point-to-point communications, happening at the same time.
- A kernel execution is a BSP superstep and is serialized with follow-up kernel iterations, or supersteps. Each stream is treated as an individual sequence of BSP supersteps during the execution. Stream interactions can be recreated in the post-execution analysis.
- Communication across BSP steps, honouring the guarantees CUDA's execution model gives, is not bound to any timing or ordering during the execution of a single compute kernel. Negative performance impacts by the instrumentation are therefore tolerated, because in a well-written program the negative impact does not impede correctness.

Chapter 5

Dynamic Instrumentation Using Code Analysis And Transformation For Detailed GPU Communication Analysis

This chapter describes how the instrumentation to generate trace data is realized. Goal of the instrumentation is to modify the original source so it generates data at run time that can be used to identify communication between CTAs in an application. As global memory is the only address space allowing side-effects beyond BSP supersteps, global memory operations are the target of the instrumentation. Data is generated by creating a trace record every time a global memory operation is performed. Before a detailed explanation of the steps, an overview of the involved components is given. The instrumentation happens during the build of the program and is handled in Clang and LLVM. They are used to insert code and functions for data generation.

As libraries present external code not in source or LLVM IR, code using libraries can't be reliably instrumented.

5.1 Trace Setup

Figure 5.1 shows at which points the original application is extended. The instrumented kernel generates data, which is consumed by a thread running on the host. This consumer thread forwards the data to a storage, in this case a simple file. Stream callbacks are used to encapsulate the kernel between two functions controlling the trace consumer. This guarantees isolated traces for every kernel in a stream. CUDA streams allow concurrent execution of kernels and DMA data movement on one device. A stream also allows executing multiple kernels at the same

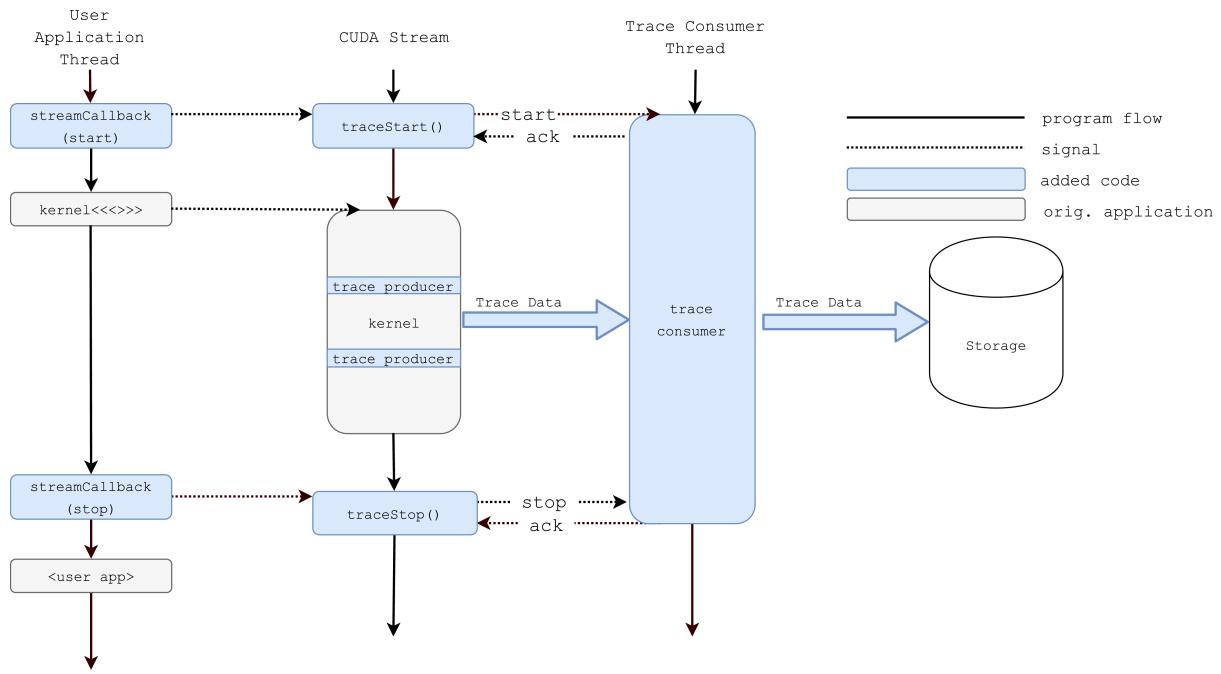


Figure 5.1: Structure of an instrumented application. Segments added during instrumentation are highlighted in blue.

time, as long as resources are available. Multiple concurrent kernels, using the same buffer lead to unordered data, that is not clearly assigned to a stream or kernel. Post-morten separation of data is infeasible. Multiple consumer threads, each managing one stream and private buffers and storage for each of them resolve this issue. Interaction between streams can be tracked by cross-referencing the stream files in the later analysis of the data.

5.1.1 Build Process

Host side modification happens in a Clang front-end plugin. This plugin extends kernel and function declarations and definitions and includes functions used by the trace. This generates a new file that is used instead of the original source . The instrumentation to trace memory operations in the kernel is done by an LLVM optimization pass and happens during the compilation of the file augmented by Clang. After compilation and instrumentation, the created object files are linked together with the rest of the application and object files used by the trace. Clangs task is detailed in section 5.2, LLVMs process is elaborated in section 5.3 and section 5.4 covers details on how data is buffered and transferred from device to host at kernel run time.

```

1 | 32 bit | 4 bit    | 28 bit | 64 bit | 64 bit      |
2 | 191   | 160 | 159   | 156 | 155   | 128 | 127   | 64 | 63   | 32 | 31   | 16 | 15   | 0 |
3 |SMID   |Type    |Size     |Address|CTA.x   |CTA.y   |CTA.z   |

```

Figure 5.2: Trace record format

```

1 <size of one record, in bytes> //stored in a single byte
2 <kernel name> // name of kernel, followed by LF
3 <records> // all trace records, as one binary blob
4 00000000000000000000000000000000 // EOT
5 <kernel name>
6 <records>
7 00000000000000000000000000000000
8 <EOF>

```

Figure 5.3: Trace file format

5.1.2 Trace Format

A trace record consists of 192 bits of data as detailed in code snippet 5.2, stored in LSB order. It contains the executing SM, CTA ID, size of the data type, type of memory operation and the operation's address. The record format allows to track the communications between CTAs across BSP supersteps on a per-address basis. The format of a complete trace file is shown in listing 5.3. The first byte in a file is the record length, followed by a LF. The next line contains the traced kernel's name. Because all records have the same size, they are stored without a separator. End of trace (EOT) for a kernel is marked by a record consisting of only zeroes. The next line contains the name of the follow-up kernel.

5.2 Source to Source Compilation in Clang

The first instrumentation step extends the host code and adds includes that are used by the kernel. A Clang front-end plugin is used to transform a file and generate a new, augmented source for the build process. The plugin uses the `ASTVisitor` pattern in clang's API to traverse statements and declarations of interest. After the AST modification, two headers are included and a new file is created.

- `TraceUtils.h` contains host-side buffer and stream management, trace consumer control and storage.
- `DeviceUtils.h` contains `__device__` functions used for the trace. The file is in-

cluded in all files containing kernel code because the following LLVM optimization pass will create calls to these functions in the original kernel.

AST Declarations contain kernel and `__device__` function declarations. They are extended with the following arguments:

- **Reservation Index Array:** Array of indices for buffer space reservation. Because each buffer is segmented into various bins, each bin needs its own index.
- **Acknowledgement Index Array:** Array of indices for write acknowledgements. Because each buffer is segmented into various bins, each bin needs its own index.
- **Buffer Pointer:** Base pointer of the buffer.
- **Bin Size:** Max. number of elements per bin.
- **Bin Count:** Number of available bins.

The mentioned bins are a logical division of buffer space due to reasons explained in 5.4. They are not be mistaken for the private buffers of each CUDA stream! Each stream buffer is divided into bins.

AST Statements contain the calls kernels and `__device__` functions. The augmentation of the latter is trivial. The kernel's trace arguments are passed to the function. The former however needs more manipulation.

First, the CUDA stream is extracted from the kernel parameters. Two statements are inserted before the kernel call. The first one allocates stream buffers, creates the storage and spawns the consumer thread in a waiting state. This only happens once per stream. Next, the function starting the consumer thread is issued to the stream, using `cudaStreamAddCallback()`. Now the kernel call is extended using the buffers allocated for this stream. After the call the function stopping the trace consumer is issued to the stream. It guarantees that the buffered data is consumed and stored before starting another trace. An explicit stream synchronization is therefore not necessary.

5.3 Code Transformation in LLVM

Instrumenting the application to generate trace data for global memory operations on the GPU happens inside an LLVM IR optimisation pass. The original source is analysed and instrumented

```

1 %arrayinx1 = getelementptr inbounds (float, float* addrspacecast (float
2     addrspace(3)* @_ZZ5saxpyfPfs_iPlS0_E2_x to float*), i64 0, i64 2)
3 store float 1.000000e+00, float* %arrayinx1

```

Figure 5.4: Example for address space lowering in LLVM IR

at the locations of loads, stores and atomic operations targeting global memory, so that during the execution data is generated.

Global memory is the only address space of CUDA in which side effects that are visible across kernel completion boundaries are possible. These side effects can carry information from a kernel to another and so, only memory operations using this address space need instrumentation. Finding loads and stores targeting global memory is not trivial in LLVM IR. Clang performs lowering of all memory access operations during the generation of LLVM IR and in the process eliminates the address space information at the location of the memory operation. An example for the lowering is displayed in listing 5.4. The `getelementptr` instruction is used for address calculation, for example to access a specific element in an array. However, before this an `addrspacecast` casts the pointer from address space 3 (shared memory) to address space 0 (generic address space). The pointer used in the `store` instruction in line two does not contain information about its address space. This lack of information means the LLVM pass can not visit loads and stores exclusively to determine the address space, and a more complex approach is necessary.

PTX does not necessarily need address space information for a memory operation since it can be resolved dynamically during execution. While PTX has distinct load and store instruction for each address space, a generic version is also available. This generic instruction maps to global memory, if it does not fit into a window for the other address spaces [?]. This lookup however comes at the cost of a minor performance decrease [?]. So, while beneficial it is not necessary to have the address space information to build a correct application. For the tracing process, however, we need to have this information to prevent the generation of trace data for parts of the address space that are not required for this work.

There is LLVM optimizer which tries to infer the address space by analysing the steps that have been introduced by the lowering [?]. However, it works on a function scope which is not sufficient for our purposes. Address space information doesn't exist in `__device__` function arguments, because CUDA allows any kind of address space in pointer type arguments. And due to the pass' function scope, information on the calling context is not available. Because the pass can fall back to the generic address space, it is not critical if the address space cannot be determined reliably. This is not sufficient for the needs of this work because the address space in-

formation needs to be definite. Therefore, an analysis generating definitive information, whether a memory access targets global memory or not, is performed. If generating this information is not possible, the pass fails. The reason for possible address space ambiguity is explained later.

LLVM widest analysis scope is 'Module', which roughly corresponds to one translation unit in C. Interactions of the original source that reach beyond this Module scope cannot be analysed, and therefore not instrumented.

5.3.1 Address Space Analysis

The basic principle of this analysis rests on the limited number of sources in a kernel, where a global address space pointer can be obtained from.

- direkt kernel parameters
- indirekt kernel parameters (struct etc.)
- global variables
- pointer type load from global memory

Collecting base pointers form all these sources provides starting points for the analysis of the program. From the def-use chain of each pointer and subsequently following each users new definitions, a usage tree for each of base pointers is created. And only memory operations that are part of def-use trees which root in the global memory pointers, can be a candidate for instrumentation. The used algorithm performs its analysis on the sparse program representation of def-use chains, created by the SSA form of LLVM's IR. The algorithm follows the order of instructions through the program, because only with proper instruction order we can obtain the address space information correctly. Therefore, the analysis is 'flow sensitive'. The location and arguments of function calls matters for our analysis, for reasons later in this section. This classifies our algorithm as 'context sensitive'. The property space of the analysis describes sets of instructions (...*Inst*) in the context of global memory operations that are defined as following:

Definition 5.1 $MOp = \{LoadInst, StoreInst, CallInst\}$

MOp contains all types of instructions interesting for tracing. During the analysis they are marked for later instrumentation. LLVM maps CUDA's atomic operations to functions, which is why *CallInst* is part of this set. They are only considered a memory access, if the callee is an atomic function.

```

Data: original pointer
Result: List of global memory operations based on original pointer
add original pointer to workstack;
while workstack not empty do
    get element form workstack;
    forall users of element do
        /* apply tranfer function  $f(x)$  to user: */
        if type(user)  $\in MOp$  then
            | add user to list for tracing;
        else if type(user)  $\in POp$  then
            | add new parent to workstack;
        else if type(user)  $\in NOp$  then
            | continue with next user;
    end
end

```

Algorithm 1: How to find global memory operations based on a input pointer

Definition 5.2 $POp = \{CallInst, GEPInst, CallInst, PHINode, SelectInst\}$

POp contains operations that define new SSA variables which can be used in operations interesting for tracing. $CastInst$ because of address space conversions and $GEPInst$ since it is the generic way in LLVM to perform address arithmetic. Their users need to be visited later. $PHINode$ and $SelectInst$ are special cases. Only instructions that resolve pointers and all arguments can be related to a global memory base pointers are a part of this set. The reason for this is explained later on. This set also includes function calls. However, in the case of a function not the return value definition, but the argument passed into the function marked for later analysis. The analysis continues in the function by moving the argument onto the workstack.

Definition 5.3 $NMOp = \{inst : inst \notin POp, inst \notin MOp\}$

NOp contains all instructions that are of no interest for the analysis and are ignored.

To obtain a list of all instruction which require instrumentation, each of the orginal base pointers is analysed by Algorithm 1. It creates a list of all descendant memory operations based on the original pointers by applying the transfer function $f(x)$ to every instruction processed. $f(x)$ classifies each instruction, assigning it to one of the three sets MOp , POp or NOp . Instructions assigned to MOp are saved for later tracing, POp instructions are put on a workstack for further analysis and $MOps$ are ignored. After this, a list with the instructions for tracing of each function has been generated. This algorithm only proves if a memory operation uses a pointer originating from a original base pointer, and therefore uses the same address space, which can be achieved without the need to analyse cycles in the CFG. As a consequence the algorithm always terminates

once each relevant instruction is visited. By keeping track of each function that was already visited, recursion is not an issue during the analysis.

ϕ -Node Limitations

Static analysis has limitations that can cause an unsuccessful address space analysis. This particular issue is caused by the SSA's need to define a new version for a variable each time said variable changes. This results in the need of ϕ -functions and `select` statements, which is a LLVM IR instruction to handle small `if-else` branches without the need for multiple basic blocks and a ϕ -node. Figure 5.5 shows a minimal code example that can lead to a situation which is not analysable by our static analysis, because it requires resolving pointers to different address spaces at runtime. The code sample on the left creates a situation where a ϕ -node is required to resolve the edges of the incoming basic blocks. Each of these basic blocks defines a surrogate pointer for `d`. For this example we will call them `ds1` and `ds2`. Based on `ds1` and `ds2` the ϕ -node defines a 'final' pointer `df` that is used for the following operations performed on `d`. Therefore the ϕ -node resolves which address space `df` points to.

This is illustrated by graphs on the right hand side of the figure. Every block preceding the ϕ -node carries an address space, represented by a colour. The left graph has to resolve different address spaces and therefore the address space for `df` is ambiguous. This leads to an unsuccessful analysis. The right graph shows a special case the implemented pass can resolve by proving that `ds1` and `ds2` both point to the same address space, which concludes that `df`'s address space is definite. The proof is implemented as a table, counting the number of times the analysis pass visits every phi-node. After the analysis is finished the value in the table is compared to the number of the ϕ -node's incoming edges. If the counted value and number of edges match the pass proved the special case, otherwise we have ambiguity in the address space of `df`, which leads to aborting the instrumentation pass.

Ambiguous Function Address Space

Another issue is possible ambiguity in the address space of pointers that are passed to functions. To illustrate this issue, on the left hand side figure 5.6 provides a code sample of a coalescing copy function. CUDA allows calling this device function with pointers to every address space and every combination of address spaces. The kernel on the right hand side then uses the function twice, once to copy data from the global memory into shared memory and a second time to move the data back to the global memory.

In a naive approach, instrumentation pass analyses the kernel, finds the first function call and the

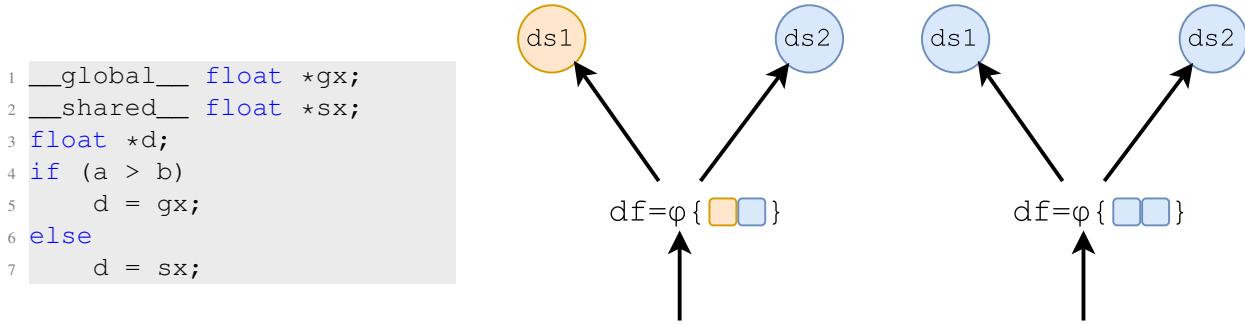


Figure 5.5: Example for static analysis limitations. Different colors represent different address spaces. The left graph represents the generic problem case and right graph a special case that is resolvable by the analysis.

first parameter is a global memory pointer. It enters the function and marks the `load` instruction for tracing. The pass continues and later finds a call to the same function, enters the function again and this time marks the `store` for tracing, because the pointer led the pass to it. Now both memory operations are marked as global memory operations and generate trace data, although they are not always used with a global address space pointer. This issue is the reason, our analysis and the added instrumentation needs to be context sensitive. Potentially, each calling context has unique pattern for tracing.

Context sensitive tracing is implemented using function specialization (or procedure cloning). This work uses a variation of the algorithm introduced in [?]. During the analysis a table keeps track of each function call, and which of its arguments belongs to global memory addresses. While gradually working through the initial list of global memory pointers and tracking their descendants during the analysis, a function call might be visited multiple times and each time a different argument (and thus different pointer) led the analysis there. The recurring visits form an argument pattern for each function call. After the analysis, all function calls are reduced to a pattern list for each function. Every pattern contains a combination of arguments that belongs to the global memory address space. A pattern is stored as an unordered tuple with the length corresponding to number of global memory space pointers and each element is the index in the function's argument list. For our example the analysis creates two patterns, **{0}**, **{1}**. A third option would be **{0,1}**, which would indicate both pointers belong to the global memory address space. There can be no empty patterns because the analysis would never visit a function call where no global memory pointer is involved.

Now that we know which patterns exists, the original function is cloned once for each pattern. Therefore, the example generates two clones, `copy_0` and `copy_1`. Every created clone is assigned to one function call pattern. Following that, every function call in the created table gets his callee replaced with clone matching the pattern of the call. Based on the call's arguments the

```

1 __global__ void k (float* gx) {
2     __shared__ float* sx;
3     copy(gx, sx);
4     // kernel does stuff
5     copy(sx, gx);
6     return;
7 }
```

Figure 5.6: Example for ambiguous function address spaces

Data: Function Call Table

Result: List of global memory operations in cloned functions

forall *function calls in table do*

- get argument pattern of call;
- get or create clone for pattern;
- replace callee of call with clone;
- get new argument entrance pointers;

end

forall *new clone pointers do*

- re-analyse all new argument pointers;

end

Algorithm 2: Algorithm for function specialization

analysis is executed again for each function call, now marking the clone's instruction for tracing. During this process, the original function's instructions are removed from the tracing list.

After this process has finished, multiple versions of a function have been created and each of them only creates trace data for certain memory operations. Specifically, for the example two functions would have been created, one that augments the memory operations on `src` and one for `dst`. The respective calls to `copy` would have been replaced with `copy_0` and `copy_1`. It is possible that more than one parameter is used for tracing.

5.3.2 Instrumentation

The optimizations pass' last step is the actual instrumentation. Some preparation is performed once in every function, before the actual tracing instrumentation takes place.

1. Fetch the arguments added by the source-to-source compilation. These arguments contain the trace buffer information.
2. Add instructions to fetch SMID during runtime. The call

```
1 asm("mov.u32 %0, %smid;" : "=r"(sm) )
```

copies the value from the special read-only `%smid` register into the register that is accessed by variable `sm`.

3. Add instructions to calculate trace buffer bin from arguments:

```
1 nSlots = 1 << PowerOf2OfNumberOfSlots // 1000
2 slotMask = nSlots - 1 // 0111
3 bin = slotMask & sm
```

After the preparation, every instruction marked for tracing in this function is instrumented for tracing by adding a call to the producer function. The producer function was added by the Clang front-end plugin. The call is added before the actual memory operation.

1. Get `Type` descriptor of memory operation. `Type` specifies the kind of memory operation specifically and is defined in table 5.1
2. Fetch pointer operand that is relevant for tracing
3. Calculate the size of the memory operation in bytes. Complex data types like `structs` are recursively explored and the total size of the element is calculated.
4. Insert instructions to generate left-most part of a trace record:

```
1 64bitDesc = sm << 31
2 64bitDesc |= Type
3 64bitDesc |= SizeInBytes
```

5. Insert call to trace producer function right before the actual memory operation. This function builds the last part of the trace record, describing the CTA and then writes the data into the trace buffer selected by `bin`. The details of how the data is written into the buffer are explained in section 5.4.1.

This is performed for every function that requires tracing. After the pass finishes, the application can be built and linked.

5.4 Tracing Process

Many applications have non-deterministic executions, for example no predetermined number of kernel calls because of varying input data. Examples are graph algorithms depending on implementation and data, or numerical solvers iterating until a convergence condition is reached

Memory Operation	Descriptor
Load	1 « 28
Store	2 « 28
AtomicAdd	3 « 28
AtomicSub	4 « 28
AtomicExch	5 « 28
AtomicMin	6 « 28
AtomicMax	7 « 28
AtomicInc	8 « 28
AtomicDec	9 « 28
AtomicCAS	10 « 28
AtomicAnd	11 « 28
AtomicOr	12 « 28
AtomicXor	13 « 28

Table 5.1: Memory operation descriptor table

by the residual. Because of this, a static buffer that is filled up once and is cleared at the end of the application does no suffice for reliable and dynamic tracing.

Therefore, a producer-consumer managed buffer for the generated trace data between device and host is used, ensuring the application never runs out of buffer space for the tracing data. The GPU generates the data and the host evacuates full buffers, storing the data. The buffer containing the generated data is host-mapped memory, which is separated into several bins. The bins are used to reduce pressure on the memory system that is generated by the atomic index accesses of the producer-consumer setup. The number of bins is always a 2^n and a scheduled CTA uses the n least significant bits of the SM's ID it is scheduled on, to select a bin. Therefore the number of bins should be selected as the closest power of two, to the number of SMs.

Writing the data to disk after they have been cleared from the buffer by the consumer, showed to be a performance critical aspect during the trace. Per default, Linux directs `writes()` to a page cache. Smaller, periodic writes (64Kbyte to 1Mybte) to the disk empirically showed to be more efficient. The author assumes this is caused by overlapping between file cache management by the OS and queue management.

5.4.1 On-Device Producer

This section focusses on the producer that is running on the device, making sure that data is only written if there space left in a buffer. The producer-consumer setup uses a head and a tail index to reserve buffer space and write acknowledgements to signal that the write has completed. Code sample 5.8 shows the the naive algorithm that is performed by the producer. First the required

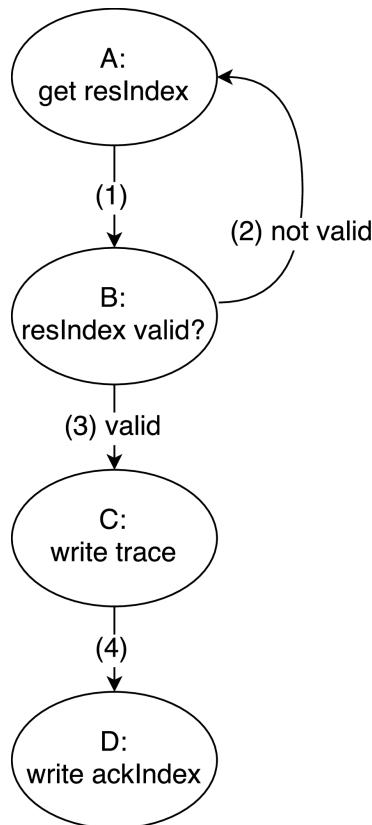


Figure 5.7: Producer CFG

```

1 while((ackIndex > maxIndex) or (id = atomicAdd(resIndex, increment) >
    maxIndex));
2 buffer[id] = traceData;
3 atomicAdd(ackIndex, increment);
  
```

Figure 5.8: Device Producer, naive approach

buffer space is reserved for writing by incrementing the reservation index by the number of active threads (line 1). Then the data generated by the trace is written to the buffer (line 2) and the write acknowledgement is incremented by the same number of buffer slots (line 3). The naive approach however, is prone to deadlocks.

The used implementation is operating on warp scope, which means either the whole warp continues or waits, to prevents deadlocks during indices fetching. A deadlock can occur because of how GPUs manage branch divergence inside a warp. If the instruction stream of a warp branches, all member threads of this warp execute both branches, and the group of members that should not be executing this part of the code is masked out. After all branches completed execution, the combined execution of both groups continues.

A deadlock can occur, if each thread fetches it's `resIndex` individually and there is not

```

1 active    = __ballot(1); // get bitmap of active threads
2 nActive   = __popc(active); // get count of active threads
3 lowest    = __ffs(active)-1; // check which thread has to perform sync
4 rlane_id = __rlaneid(active, lane_id); // each thread gets its relative Lane
   id, based on number of active threads
5
6 if (lane_id == lowest)
7     while( ackIndex >= maxIndex-maxWarpWrite
8         || (id = atomicAdd(resIndex, nActive)) >= maxIndex-maxWarpWrite
9     );
10 // Warp stalls here until all branch is completed
11
12 idx = __shfl(id, lowest) + rlane_id; // distribute id
13 buffer[id] = traceData;
14 if (lane_id == lowest )
15     atomicAdd(ackIndex, n_active);

```

Figure 5.9: Device Producer, warp scope

enough available buffer space for all threads inside the warp. Referring to figure 5.7 now, all threads will execute edge (1) leading from (A) to (B) by fetching `resIndex`. If all threads get a valid index (B), the warp will not branch and the trace can complete successfully. If the check in (B) fails for one or more threads in the warp, a branch is created. The branch without valid IDs will use edge (2), returning to node (A) and retry fetching a valid `resIndex`. The branch with valid `resIndex` has to wait with progressing on edge (3) until the group without valid `resIndex` is also ready to continue on edge (3). But because the warp can not progress over edge (3) to C and from there over (4) to D, until all threads have a valid `resIndex`, the threads in the warp that already have a valid `resIndex` can not reach node D write `ackIndex`, which would lead to a evacuation of the buffer on by the host and a reset of the `ackIndex` and `resIndex`. As a result of this, a deadlock occurs because one group of threads waits for a resource that can not be released by the other group of threads due to the warp execution model.

This is resolved by handling reservation and write acknowledgement on a warp level. One thread makes the reservation and write the acknowledgement for all threads inside the warp. Now either the whole warp can finish the trace or the whole warp stalls. This way we can use the warp execution model for our purposes, because one thread that waits for the IDs can stall the whole warp with an if branch and no other means of synchronization. This is implemented by using CUDA warp intrinsics, which offer efficient communication inside of warps. Little extra code is needed shift everything to a warp-scope management, as seen in Figure 5.9. Because it is possible that a trace occurs on a point during the execution, where the warp is already branched we have to make sure that the indexes are managed correctly and only the currently active threads participate in the trace. The warp intrinsics `__ballot`, `__popc` and `__ffs`

are used to determine how many threads are active, and which one has to perform the atomic operations to handle the producer synchronization. Then `__rlaneid()`, which is user code, calculates the relative lane id for each warp.

While the lowest thread fetches the ID for the buffer using `atomicAdd`, the whole warp stalls at the end of the if clause because of the warp execution model. Once a valid ID is fetched, it is distributed among all threads using the `__shfl` intrinsic. Then the trace is written and the lowest thread increments the write acknowledgement by the same number it increased the reservation.

Notice that in this case, the conditions for fetching the id now takes into account that data is written in blocks and has to check if there is still space left for one complete block.

Assuming a warp is fully populated at the moment of a trace, and thus always reserves enough buffer space for a whole warp, can create gaps in the trace data. To prevent old data from lingering, the buffer would need to be wiped after every evacuation of the generated trace data. Additionally, these gaps would require additional handling in the analysis of the data. Therefore, this more complex approach of handling the consumer was chosen, which only reserved as much space as required.

5.4.2 Host Consumer

The consuming thread on the host side is much simpler. A single thread is iterating over all bins in the buffer, checking if the `ackIndex` is at the threshold for evacuation. This threshold is reached if there is not enough space for a complete warp to write a trace. The buffer is evacuated by writing the data into a file handle and `resIndex` and `ackIndex` reset to the beginning of the buffer.

Chapter 6

Evaluation and Analysis of GPU Communication

This chapter analyses the traces generated with the technique described in chapter 5. The analysis is sectioned into several tiers, as described in 4.2.1.

6.1 Evaluated Applications

The evaluated application are listed in table 6.1. The applications are selected to represent typical problem domains in HPC. *Hist*, *hs2d*, *hs3d* and *nbody* are regular applications, where each CTA

Application	Source	Domain	Abbreviation	Problem Size
Hotspot 2D	Rodinia	Structured Grid 2D	hs2d	512x512, t = 10
Hotspot 3D	Rodinia	Structured Grid 3D	hs3d	512x512x8, t = 10
Histogram64	CUDA SDK	Histogram	hist	64MB into 64 Bins
NBody	ZiTi CEG	Nbody	nbody	512 Bodies, t = 5
Pathfinder	Rodinia	Shortest Path	path	100000x100x20
Breadth First Seach	Rodinia	Graph Processing	bfs	10^6 Nodes

Table 6.1: Application used for communication evaluation

only operates within his data region. *BFS* and *path* contain irregularity, as the execution is dynamic and the behaviour is depended on the input data.

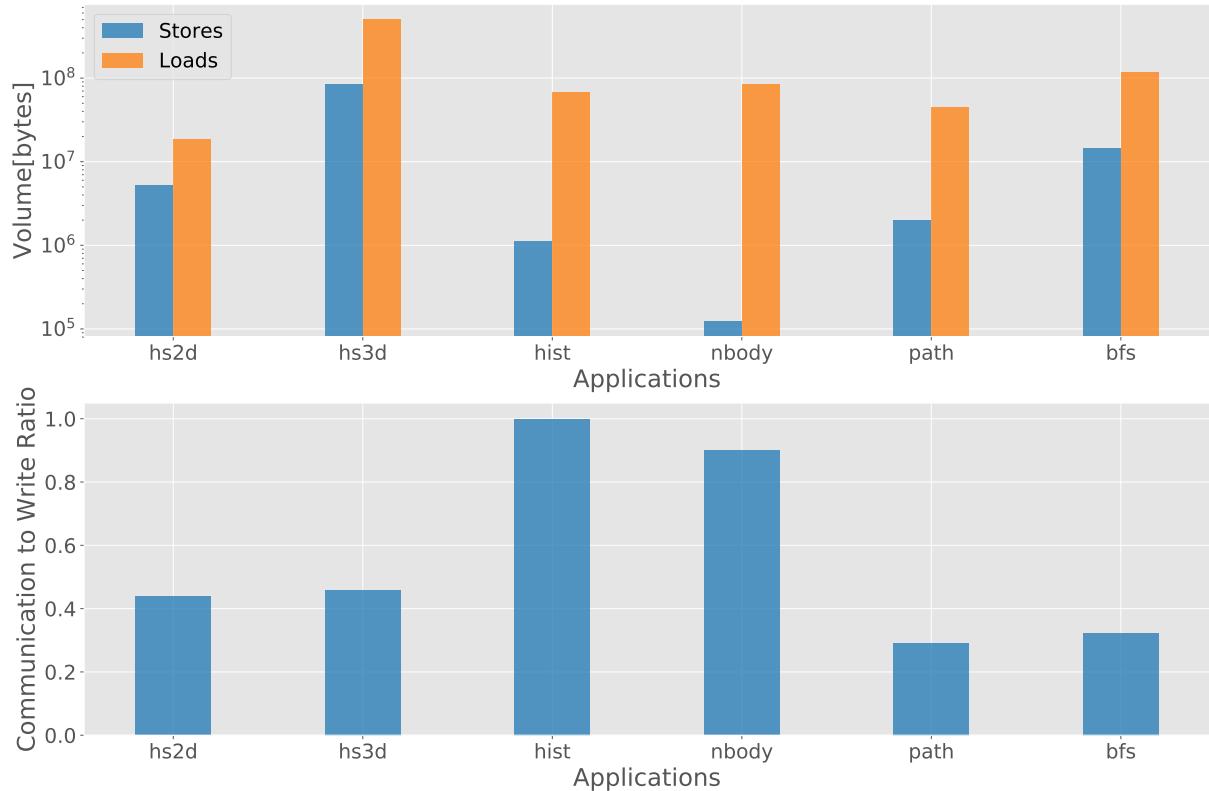


Figure 6.1: Total data volume and ratio of communication to total of all writes.

6.2 Tier I Analysis

6.2.1 Total Data Volume and Communication Fraction

Figure 6.1 directly compares the total data volume of loads and stores of and how much of the write volume is involved in communication. At least 30% of any application's writes is accessed by another CTA or another kernel in following iterations. As expected, loads exceed the stores by at least a single order of magnitude.

Nbody and *hist* show a very high ratio of communication, because *nbody* is an n -to- n problem and the histogram algorithm reduces a problem into bins, only using on the data generated in the preceding step. The irregular applications, *bfs* and *path*, communicate less data compared to the four regular applications.

Against initial expectations, *hs2d* and *hs3d* show almost 50% communication ration. The reason is, *hs2d* uses deep halos which increase the number of transferred bytes between iterations. *Hs3d* exchanges the complete z -dimension across iterations.

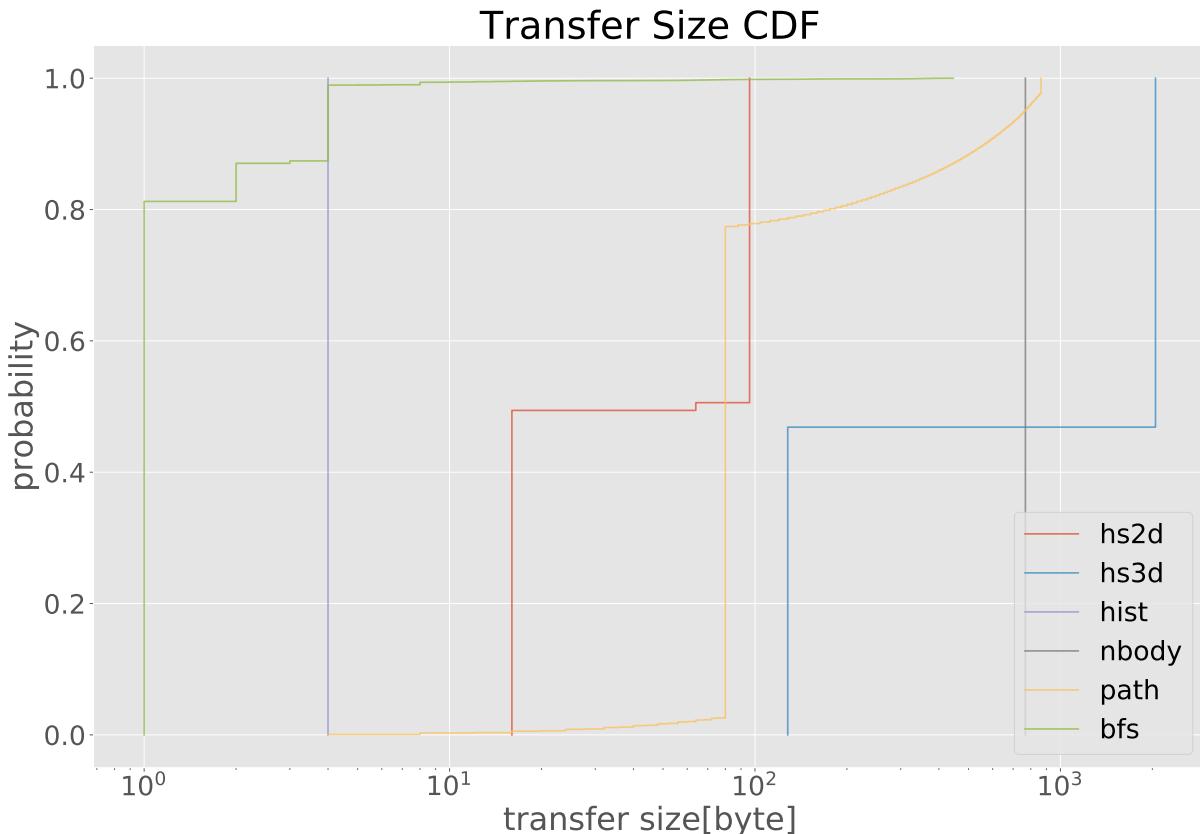


Figure 6.2: Message Size CDF Plots

6.2.2 Transfer Size CDF

Figure 6.2 shows the cumulative distribution functions (CDF) plots for each application. The CDF shows the cumulative probability of a transfer size between two CTAs to occur during the execution. All applications, except *path*, show one or few dominating transfer sizes. The *path* application shows two slopes with linear rising probability for a range of transfer sizes.

The dominant transfer size of *bfs* is caused by the frontier array, used to communicate across graph segments. The *hist* has one transfer size, because only one uniform kind of transfer happens in this application. This is also true for *nbody*. The steps in *hs2d* and *hs3d* are caused by the problem space borders in the calculation

6.2.3 CTA In/Out-Degree

The in and out degree describe the number communication partners, separated between incoming and outgoing transfers, for each CTA. Figure 6.3 plots the percentages of selected bins for each application.

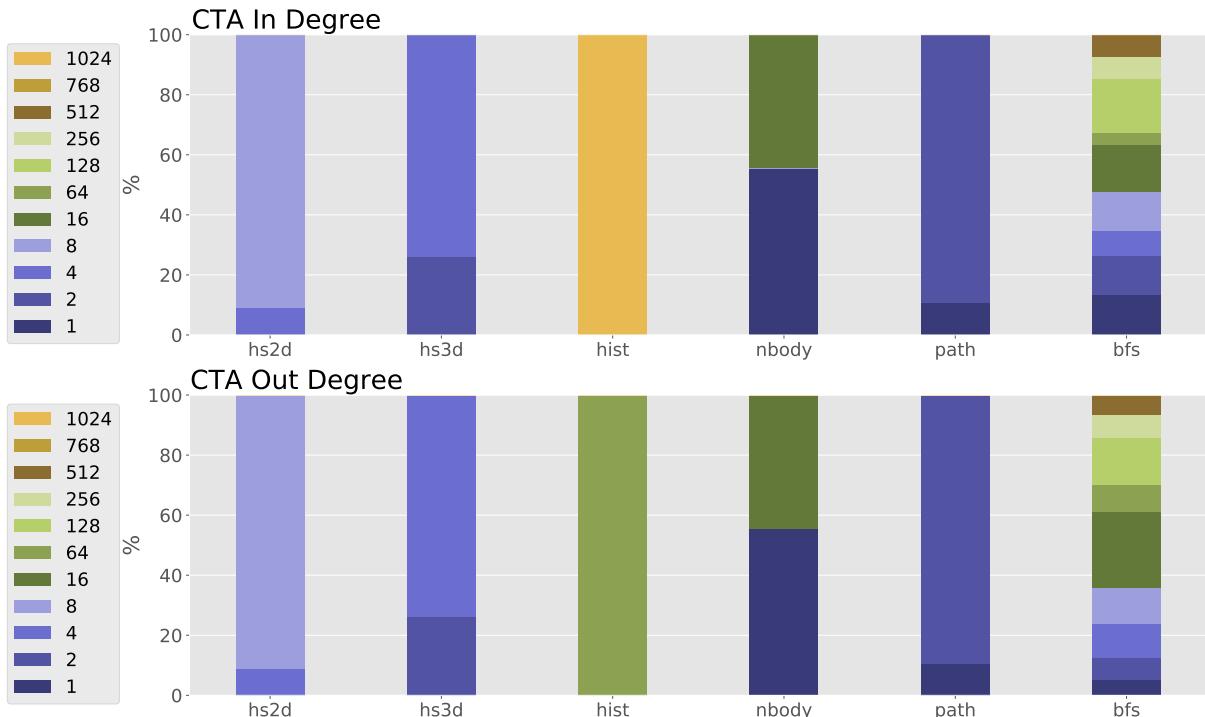


Figure 6.3: Number of incoming and outgoing communication partners of each CTA

Except for *nbody*, and to a smaller degree *bfs* and *path*, no application performs point-to-point communication. Almost every communication happening is, to some degree, a collective communication.

Most applications show low variance of in- and out-degree, and few communication partners. *Bfs* has a high variation because the traversed graphs dictates the communication partners. The partners of *hs2d* are a result of 2D tiling and *hs3ds* partners dictated by stencil size. Also notable is that although *path* is a dynamic application, the number of communication partners is steady. The in/out-degree of *hist* stems from the two-step process, where each CTA in the second step reads a value from every CA in the first step.

6.2.4 Bisection Volume

This metric shows the data volume moved a hypothetical cut in each dimension. The cut creates two equally sized grids. This can help analyse, how well an application scales to multiple GPUs. Because none of analysed application use the *z*-dimension of CTA organisation, only the *x* and *y* dimensions are plotted.

Hs3d strikes out, as it clearly shows that a cut in the *y*-dimension would not be favourable. The *path* algorithm shows a very low bisection volume, making it a good candidate for multi-

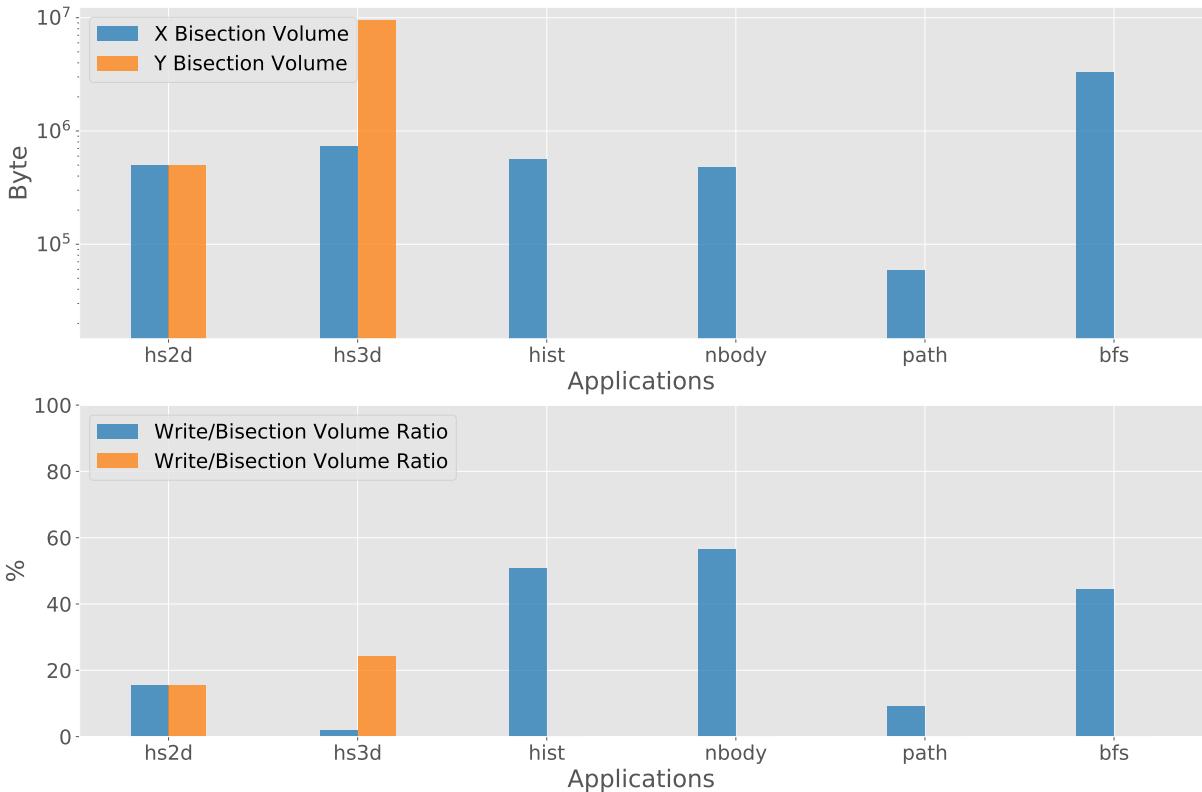


Figure 6.4: Total and relative bisection Volume, separated by dimensions

gpu scaling. Because of *bfs* very irregular access patterns and high bisection volume and ratio, scaling to multiple GPUs can be challenging. The same goes for *hist* and *nbody*, because they both perform all-to all communication. The scalability of the stencil applications depends on the problem, stencil and halo size and halo.

6.2.5 Communication Stride

The stride in loads and stores involved in communication can be a critical aspect for performance. Large strides reduce the available memory bandwidth and increase latency. The strides in figure 6.5 are measured in byte distance, not elements.

The *hist* application shows a large load stride in figure 6.5, together with a large in-degree in figure 6.3 suggest a 'gather'-like characteristic. For *hs2d/hs3d* we see the expected stride created by the tiling/stencil, accessing multiple rows at once. The irregular pattern of *bfs* is caused by the frontier array, which is used to communicate between all CTAs. The stride of *nbody* is



Figure 6.5: Strides of loads and stores used in communication, by CTA

6.2.6 Density Map

The density plots, sometimes called heatmap, describes the frequency of CTA interaction across BSP supersteps. The x-axis are the CTAs performing writes, the y-axis are the follow-up reads to each write. Different kernels in an application are ignored, and only the CTA ID is used to identify communication partners. Figure 6.6 displays a subset of 200 CTAs for each application. The IDs of all CTA are serialized in x,y,z order, x being the most dominant dimension.

The regular application show the expected patters. The stencils show communication with their neighbours according to stencil and tiling, *nbody* shows communication with all other CTAs, same as *hist*.

The irregular applications show two very different patters. *Path* has only few partners but the frequency varies significantly. *Bfs* shows two extremes, high frequency with high regularity and low frequency with low regularity. The first extreme is caused by two kernels alternating during the iterations, where CTAs with the same ID process the same segment of the graph. Latter is caused by the frontier array, used at the segment borders. A different graph would create the same pattern of the first extreme, but a different pattern of the second extreme.

The *nbody* plot does not show 200 CTAs, as it only used 15 CTAs.

As the CDF plot shows, all applications are dominated by few transfer sizes between CTAs, the

patterns in figure 6.6 would remain largely unchanged if the density was replaced with accumulated data volumes.

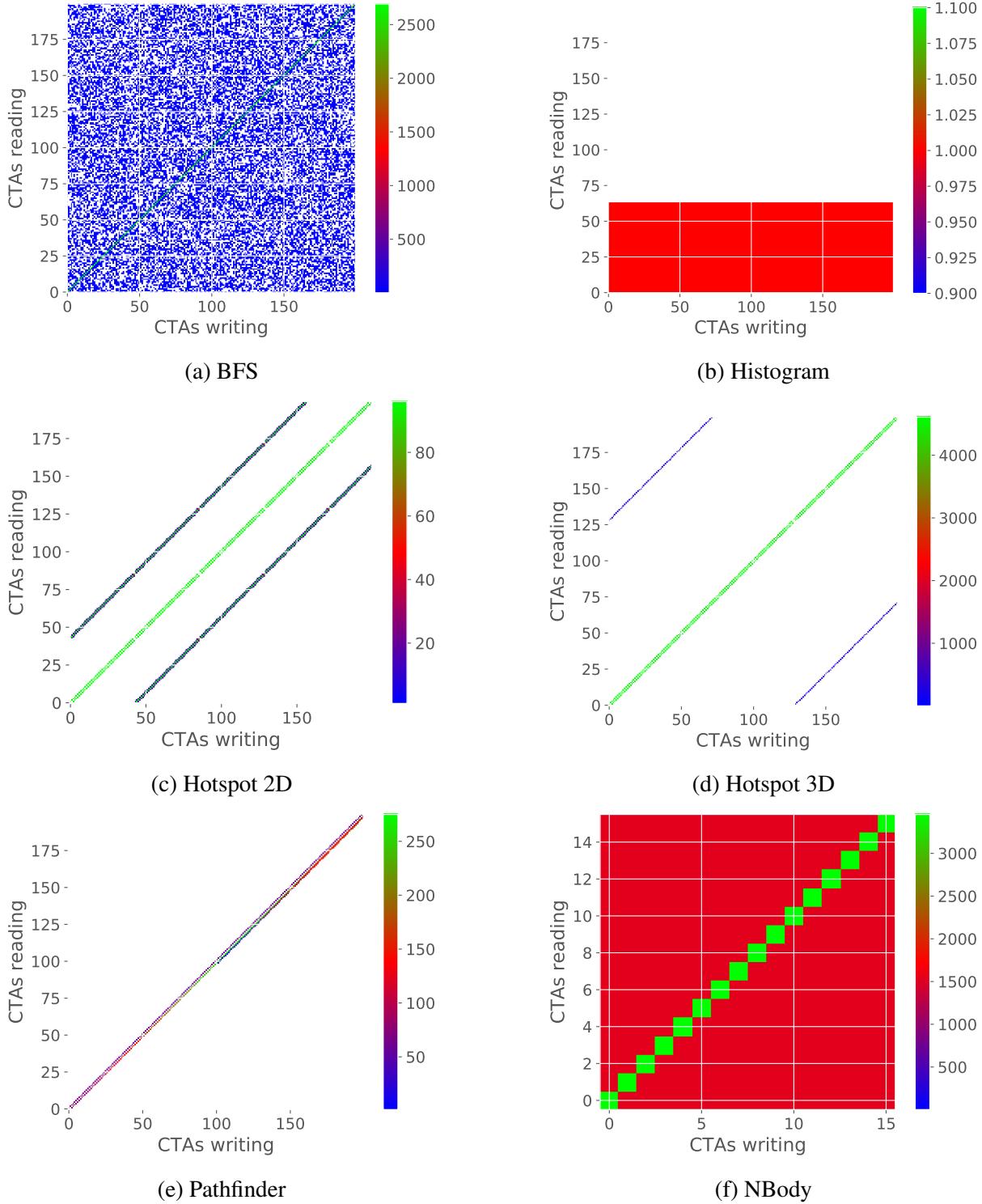


Figure 6.6: Transfer Density Plots

6.3 Tier II Analysis

6.3.1 Kernel communication Evolution

This analysis if figure 6.7 shows how the data volumes change across supersteps. It is separated into loads and stores. Loads start at superstep one because as to our definition of communication, there can be no load that is part of communication during the first superstep. Stores are not displayed in the last iteration, because the last store of an application is no communication.

The graph is to read as such: Stores happened in this superstep and are readable in the next. Loads read data from the preceding superstep. CTA in- and out degree analysis showed, every communication is happening in a collective environment. This means that there is no simple single writer, single reader matching across supersteps, which is the reason loads and stores are separated in these graphs.

As expected, the data of all regular applications follow a pattern, while the ones on irregular applications do not. The stencils do not vary volumes, because every iteration performs exactly the same task. Because each CTA only updates it's own set of bodies, *nbody* has constant store volumes. The load volume of every other iteration accesses all bodies, creating the zig-zag pattern.

6.3.2 Superstep Distance

This metric measures the volume of data transferred, but takes into account the number of supersteps between the store and the subsequent load. Only *bfs* shows this behaviour, so figure 6.8 shows only the volume for this application. The distance describes the number of supersteps, or kernel launches, happening before written data is read by a CTA. A distance of zero means, the directly succeeding kernel reads the data.

Read operations in different supersteps can reference the same write, therefore the volumes are not exclusive. As an example, a subset of distance zero might very well be included in more distant reads.

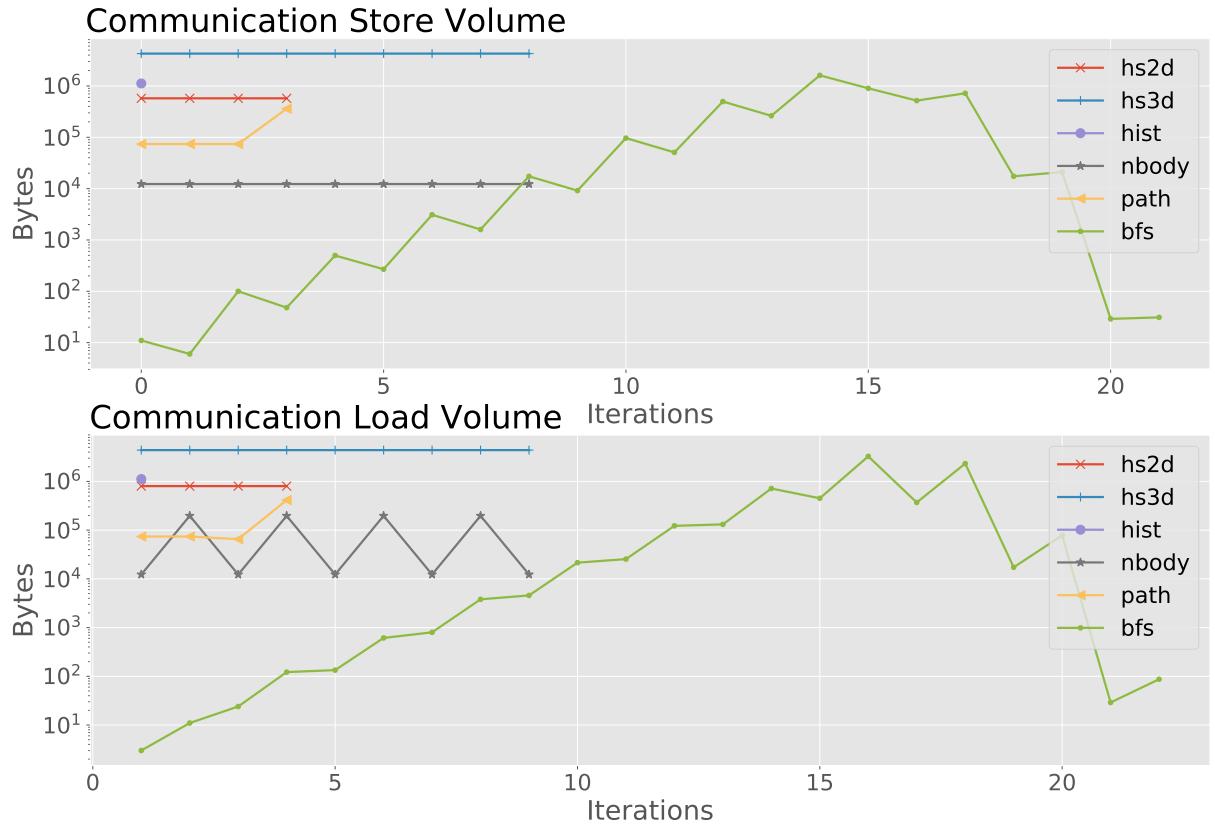


Figure 6.7: Volume of communication loads and stores in each iteration

6.4 Tier III Analysis

6.4.1 CTA connectivity evolution

As shown in figure 6.3, most applications have a steady number incoming and outgoing connections. Figure 6.9 shows how the mean number of outgoing communication partners changes over time, including the standard deviation.

The fluctuation of *nbody* can be explained with two alternating kernels, one kernel accesses only its own set of bodies, while the other kernel accesses all bodies, which increases the out-degree. Because *hist* consists of only two steps, it is represented by single value. Both *path* and *bfs* show stronger fluctuations with increasing number of iteration, caused by the exploration of the graph. Same as with the load and store volume of each iteration, the stencils show no variation.

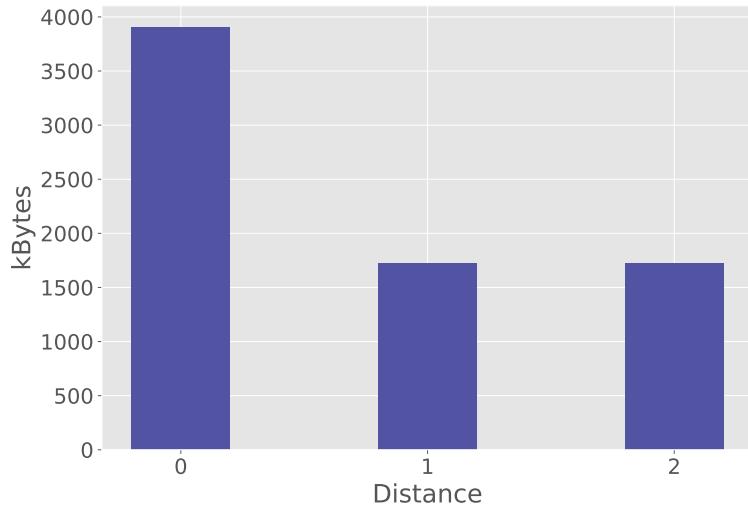


Figure 6.8: Volume of communication for *bfs* application, separated by superstep distance.

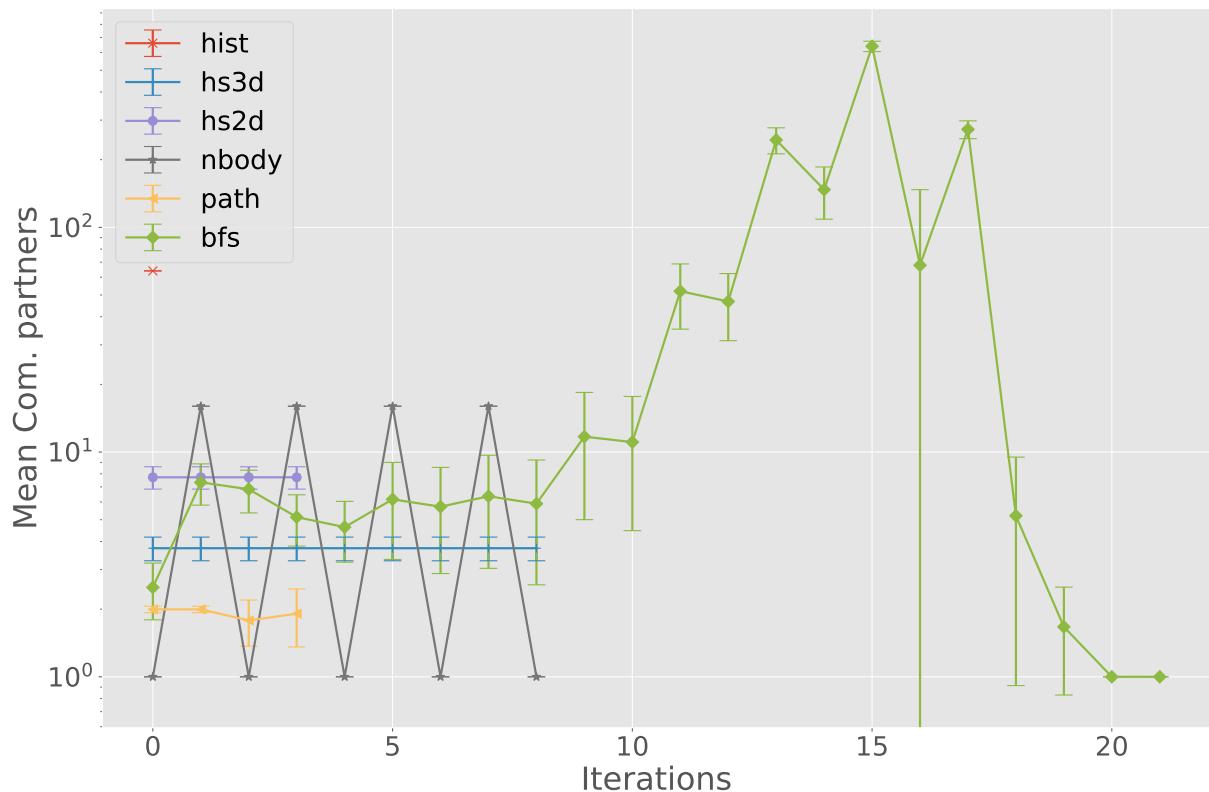


Figure 6.9: Mean out-degree of CTAs, per iteration

Chapter 7

Conclusion

The first goal of this work is using compilers to create instrumentation, making detailed communication analysis in GPU applications possible. GPUs are shared memory system and information can only be passed via global memory. Therefore, global memory operations are instrumented to create detailed memory access traces. The implemented solution introduces code on both, host and device to create a dynamic producer-consumer buffer system, allowing tracing of non-deterministic applications. Host side modifications are performed in a Clang front-end plugin, as they are rather trivial. Due to lowering and canonicalizations by the compiler, identifying global memory access instructions in device code is not straight forward and requires detailed code analysis and transformation. Analysis and transformation happen in LLVM IR, which allows a more efficient code analysis due to it's SSA properties.

Using the memory traces created by the instrumented, chapter 6 analysed the communication, as defined in 4. All applications showed that at least 30% of all writes is later consumed by a communication partner. Both, incoming and outgoing communication has collective tendencies and is rarely point-to-point. Communication is dominated by few different messages sizes in each application, usually packed tightly in memory as hinted by the low strides, which benefits coalescing. How the number of communication partners and the communicated volumes behave over time, depends on the type of applications. Regular applications with one kernel show no variance. Regular applications with multiple kernels show periodical fluctuations across different kernels. Irregular applications show growing irregularities over time. Analysing the bisection volume shows, that most applications would benefit from efficient GPU-to-GPU communication in a multi-GPU implementation, as the share of data communicated across a bisection can be substantial. Only one application showed communication that could span several supersteps.

7.1 Future Work

The work on this project can be continued on both the tooling and the analysis aspect. For the tooling, a more complex trace format that allows compressing coalesced memory access and the expression of access ranges of warp can reduce the impact on the runtime. Adding additional analysis features, like `inline asm` for ptx analysis and the support for macros like `ldg()` increase the range of applications that can be traced. To increase the number of supported applications, support of OpenCL could be implemented.

The analysis can be extended in both breadth and depth. Analysing more application can yield new insights, not yet present in the analysis. A deeper analysis of specific applications can lead to an improved understanding of a single application, revealing options for optimization.

Ultimately, a model for multi-gpu scalability could be derived from the analysis data.

Bibliography

Appendix A

Instrumentation

Instrumenting an application at compile time happens in two steps. The first step is the instrumenting all source file containing:

- Kernel Calls
- Kernel and `__device__` function definitions
- Kernel `__device__` functoin declarations

Each file needs to be parsed separately. For each file, a new file with the prefix 'augmented-' is created. This is then used for the further build process. Compile units bear some problems with kernels defined and declared in .cu files, which are then included in other files. The easiest workaround is to copy the kernel into the file including it, and then only use the including file for instrumentation and build.

The following snipped is an example how to augment a kernel file.

```
1 clang++ -Xclang -load -Xclang "clang-plugin/libMemtrace-AA.so"
2     -Xclang -plugin -Xclang cuda-aug -Xclang -plugin-arg-cuda-aug
3     -Xclang -f -Xclang -plugin-arg-cuda-aug -Xclang ./augmented-kernel.cu
4     -I$CUDASYS/samples/common/inc -I. -I../utils
5     --cuda-path=$CUDASYS -std=c++11 -E application/kernel.cu
```

Next, the host and device utils are compiled, for later linking with the application.

```
1 // Host
2 clang++ -c -L$CUDASYS/lib64 --cuda-path=$CUDASYS
3 -I$CUDASYS/samples/common/inc -O1 --cuda-gpu-arch=sm_30
4 -I$CUDASYS/include -o hutils.o --std=c++11 -I. -I../utils
5 ../utils/TraceUtils.cpp
```

APPENDIX A. INSTRUMENTATION

```
6 // Device
7 clang++ -c -L$CUDASYS/lib64 --cuda-path=$CUDASYS
8 -I$CUDASYS/samples/common/inc -O1 --cuda-gpu-arch=sm_30
9 -I$CUDASYS/include -o dutils.o --std=c++11 -I. -I../utils
10 ./utils/DeviceUtils.cu
```

Next, all the augmented kernel files are compiled at once.

```
1 clang++ -c -Xclang -load -Xclang $LLVMPLUGIN --cuda-path=$CUDASYS
2 -I$CUDASYS/samples/common/inc --cuda-gpu-arch=sm_30 -L$CUDASYS/lib64 -O1
3 -lcudart_static -m64 --std=c++11 -I. -I../utils -I<app-includes>
4 ./augmented-kernel.cu ./augmented-kernel2.cu
```

Finally, all compiled files are linked together.

```
1 clang++ --cuda-path=$CUDASYS -I$CUDASYS/samples/common/inc --cuda-gpu-arch=
  sm_30 -L$CUDASYS/lib64 -O1
2 -lcudart -ldl -lrt -L. -m64 --std=c++11 -I. -I../utils -I<app-includes>
3 -o application dutils.o hutils.o augmented-kernel.o augmented-kernel2.o
```

During the execution one or more files are generated in the /tmp folder. The files are named MemTrace-pipe-<n>, one for each stream.

Appendix B

Post Processing

Any software related to what happens after the analysis, is stored in 'memtrace-pass/post-processing'. The trace files can be parsed with the 'extract-subset.py' python3 script. From the original full trace, the communication subset is extracted. Three data structures are generated by 'extract-subset.py'. All structures are stored using pickle. The data structures are of type 'AutoDict()', which is defined in 'TraceInc.py'

1. Load/Store volumes, both total and during communication, for Kernels, CTAs and SMs. Separated by kernel and superstep.

```
1 "KCV" : { // Kernel Communication Volume
2     <kernel>: {
3         <superstep>: {
4             "Load" : <int>
5             "Store": <int>
6         }
7     }
8 }
9 "KDV" : { // Total Kernel Data Volume
10    <kernel>: {
11        <superstep>: {
12            "Load" : <int>
13            "Store": <int>
14        }
15    }
16 }
17 "CCV" : { // CTA Communication Volume
18     <kernel>: {
19         <superstep>: {
20             <CTA> : {
```

```

21             "Load" : <int>
22             "Store": <int>
23         }
24     }
25 }
26 }
27
28 "CDV" : { // Total CTA Data Volume
29     <kernel>: {
30         <superstep>: {
31             <CTA> : {
32                 "Load" : <int>
33                 "Store": <int>
34             }
35         }
36     }
37 }
38 "SCV" : { // SM Communication Volume
39     <kernel>: {
40         <superstep>: {
41             <SM> : {
42                 "Load" : <int>
43                 "Store": <int>
44             }
45         }
46     }
47 }
48
49 "SDV" : { // Total SM Data Volume
50     <kernel>: {
51         <superstep>: {
52             <SM> : {
53                 "Load" : <int>
54                 "Store": <int>
55             }
56         }
57     }
58 }
59
60

```

2. Map of transfers, ordered by kernel, CTA and superstep.

```

1 <source-kernels> : {

```

```

2      <source-cta> : {
3          <source-superstep> : {
4              <recv-kernel>: {
5                  <recv-cta>: {
6                      <recv-superstep>: {
7                          'Size' : <int>
8                          'cnt'  : <int>
9                      }
10                 }
11             }
12         }
13     }
14 }
```

3. All addresses involved in communication, with all operations performed on this address, in superstep order.

```

1 {
2     <address> : [
3         { // record
4             "kernel": <kernel>
5             "it"    : <superstep>
6             "cta"   : <CTA-ID, xyz order>
7             "addr"  : <address>
8             "smid"  : <SM id>
9             "size"   : <operation size in bytes>
10            "type"   : <type of MOp>
11        }
12    ]
13 }
```