Analysis of Validating and Verifying OpenACC Compilers 3.0 and Above

1st Aaron Jarmusch
University of Delaware
jarmusch@udel.edu

2nd Aaron Liu
University of Delaware
olympus@udel.edu

3rd Christian Munley
University of Delaware
chrismun@udel.edu

4th Daniel Horta
University of Delaware
dchorta@udel.edu

5th Vaidhyanathan Ravichandran
University of Delaware
vaidhy@udel.edu

6th Joel Denny
Oak Ridge National Laboratory
dennyje@ornl.gov

7th Kyle Friedline
University of Delaware
utimatu@udel.edu

8th Sunita Chandrasekaran
University of Delaware
schandra@udel.edu

2022 Workshop on Accelerator Programming Using Directives (WACCPD) | 978-1-6654-9019-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/WACCPD56842.2022.00006

***Abstract*—OpenACC is a high-level directive-based parallel programming model that can manage the sophistication of heterogeneity in architectures and abstract it from the users. The portability of the model across CPUs and accelerators has gained the model a wide variety of users. This means it is also crucial to analyze the reliability of the compilers' implementations. To address this challenge, the OpenACC Validation and Verification team has proposed a validation testsuite to verify the OpenACC implementations across various compilers with an infrastructure for a more streamlined execution. This paper will cover the following aspects: (a) the new developments since the last publication on the testsuite, (b) outline the use of the infrastructure, (c) discuss tests that highlight our workflow process, (d) analyze the results from executing the testsuite on various systems, and (e) outline future developments.**

***Index Terms*—Performance, Programming Model, Testsuite, Validation, Conformance**

I. Introduction

Heterogeneous systems equipped with CPUs and accelerators are becoming increasingly prevalent. To add to the heterogeneity, application developers also have to face the challenges of migrating code from one type of heterogeneous system to another. Code migration is not only tedious but also prone to introducing bugs. While there are a variety of language and models available for the application developers to choose from, such as OpenACC [1], OpenMP [2], CUDA [3], OpenCL [4], NVIDIA Thrust [5], and Kokkos [6] among there, it depends on what the developers are the most comfortable with and what would fetch them performance without losing portability, accuracy among other metrics of their applications' interest.

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan). This paper is authored by an employee(s) of the United States Government and is in the public domain. Non-exclusive copying or redistribution is allowed, provided that the article citation is given and the authors and agency are clearly identified as its source.

In this paper, we focus on OpenACC, a directive-based programming model that targets traditional X86 architectures and accelerators such as GPUs and FPGAs [7], validity of the OpenACC compiler implementations and their conformance to the standard specification; these will include features such as `atomic` or `parallel` constructs, clauses such as `if` or `copy_out`, or combinations of occurrences of clauses on constructs. Support from Commercial compilers include NVIDIA and Hewlett Packard Enterprise (HPE Cray). Currently, NVIDIA supports up to version 2.7. HPE Cray focuses on Fortran and fully supports OpenACC 2.0 with partial support up to 2.6. These implementations are being developed for not only NVIDIA GPUs but also AMD GPUs as we are beginning to see a number of systems supporting the AMD GPUs including Frontier, the first exascale system. Open Source compilers include GNU Compiler Collection (GCC) [8] with support for OpenACC 2.6 and Clacc (OpenACC support for Clang and LLVM) [9].

Academic compilers[1] include Omni Compiler from Riken/University of Tsukuba [10], OpenARC from ORNL [11] and OpenUH from SBU and UH (that are now outdated) [12].

OpenACC has been adopted for large-scale scientific applications on large-scale computing systems such as Summit at the Oak Ridge Leadership Facility and JUWELS at the Jülich Supercomputing Centre, among several others. Some of these applications include ANSYS1 [13], GAUSSIAN [14], ICON [15], COSMO 3 [16], MPAS microphysics WSM6 [17] and many more. Please refer to this tracker where OpenACC has been collecting published OpenACC papers [2].

To ensure application developers can successfully use the model for their large scale applications, it is critical that the implementations are validated and verified. The suite will not only check for conformance of the implementations to the specification but also push for consistency in functionality across implementations, which is key for the developers' success stories. It is a real challenge to debug if different

[1]More information on existing compilers can be found on the OpenACC webpage https://www.openacc.org/tools

[2]shorturl.at/mquv0

978-1-6654-9019-1/22/$31.00 ©2022 IEEE
DOI 10.1109/WACCPD56842.2022.00006

Authorized licensed use limited to: UNIVERSITY OF DELAWARE LIBRARY. Downloaded on December 05,2023 at 00:13:08 UTC from IEEE Xplore. Restrictions apply.

implementations lead to different interpretations of the specification. This can only lead to more ambiguities among both the developers and the users. By creating functionality tests and amassing the results from multiple compilers/versions to compare implementations' adherence to the OpenACC specifications, this testsuite will enable the compiler developers to improve the quality of their tools and ensure compliance of their implementations with the specifications of the language. Our previous publication on this effort [18] captured these discrepancies up to specification 2.5 (although at the time implementations did not fully cover 2.5 just yet).

The specifications have since advanced to 3.x with significant developments including features such as reduction on arrays/composites, C++ lambda support, updated base languages, `acc_memcpy_d2d`(multi-device report), support for Fortran Do Concurrent and Block constructs, `async_wait` on data regions, `acc_wait_any` and also providing support for both shared and discrete memory machines. So most, if not all, tests that this paper discusses have been written to adhere to the current specification. The paper also presents the infrastructure of the testsuite and validates most up-to-date versions of all available compilers including Clacc and HPE Cray's newer OpenACC implementations (supporting Fortran language).

The license associated with our testsuite is a dual license scheme. We are open to contributions from the community. The dual license scheme is designed in way to preserve the license used by the contributor, while the other license will be from OpenACC to ensure consistency in code versions, running code, and reporting of results. For more information on the project, how to contribute and our license please consult our website [3]. The paper makes the following contributions:

- Provide a novel testing infrastructure for C/C++ and Fortran tests
- Identify and report results on compiler bugs and runtime errors on all available and in use OpenACC compilers
- Evaluate different compilers' implementations for its conformance to the OpenACC specifications

Please refer to the project GitHub [4] for all the codes in our testsuite.

II. Overview of the Programming Model

OpenACC is a programming model that can express high-level parallelization in three levels: gang, worker, and vector. The gang level is the broadest and comprises one or more workers, can be conceptualized as a thread block, and multiple gangs can work independently of one another. A worker is the middle level of parallelism or a block of vectors within a gang; the purpose of a worker is to compute a vector. Vector is the lowest level of parallelism, executing at the thread level, working in lockstep with one another. OpenACC is designed to be a portable incremental optimization language, meaning the same code works on various architectures, parallelization can be added incrementally without changing the source code, and it is focused on providing comparable performance to lower-level paradigms. This is accomplished by abstracting the underlying hardware architecture away from the user, allowing for portability across several systems, which requires minimal to no changes in the implementation. Focusing on the use of general-purpose graphics card units (GPUs), the model has, especially in recent years, broadened the number of systems that can utilize this programming model.

To begin parallelizing, the user must tailor directive implementation based on the source code language. For C and C++, directives are prefixed by `#pragma acc`. For Fortran, they are prefixed with the sentinel `!$acc`. After the prefix, the user declares the implementation through various data and compute constructs using directives and clauses.

With proper implementation of OpenACC within a source code, the compiler handles parallelization and memory management. At compile time, the hardware architecture can be specified or detected by the compiler, and an architecture-specific optimized translation of the code is generated. At runtime, data is transferred to the device from the host as specified or implied by directives created in the source code, and the address is stored along with the host address to prevent a page error from the operating system.

Code 1: Simple Serial Addition of Arrays

```
int main(){
    int N = 1<<20;
    ***Variable Declaration***
    for (int x = 0; x < N; ++x){a[x]=10;b[x]=15}
    for (int x = 0; x < 1 << 14; ++x){
        for (int y = 0; y < N; ++y){
            c[y] = a[y] + b[y];}}
    return 0;}
```

Abstracting the hardware implementation simplifies the creation and maintenance of optimized codes by reducing the possibility of user error. As an example of the simplicity of translation of a serial source code into an OpenACC directive-based optimized code, we translate the serial code in Code 1 to both CUDA and OpenACC.

Code 2: Simple CUDA Addition of Arrays

```
__global__
void add_arrays(int n, double  *a, double *b, double *c){
    int x = blockIdx.x * blockDim.x + threadIdx.x;
    if (x < n){c[x] = a[x] + b[x];}
}
int main(){
    int N = 1<<20;
    ***Variable Declaration***
    double *device_a, *device_b, *device_c;
    for (int x = 0; x < N; ++x){a[x] = 10;b[x] = 15;}
    cudaMalloc(&device_a, N * sizeof(double));
    cudaMalloc(&device_b, N * sizeof(double));
    cudaMalloc(&device_c, N * sizeof(double));
    cudaMemcpy(device_a, a, N * sizeof(double),
         cudaMemcpyHostToDevice);
    cudaMemcpy(device_b, b, N * sizeof(double),
         cudaMemcpyHostToDevice);
    for (int x = 0; x < 1 << 14; ++x){
        add_arrays<<<65565, 256>>>(N, device_a, device_b,
             device_c);
    }
    cudaMemcpy(c, device_c, N * sizeof(double),
         cudaMemcpyDeviceToHost);
    cudaFree(device_a);
    cudaFree(device_b);
    cudaFree(device_c);
}
```

In order to translate this to CUDA, first, CUDA requires that we initialize pointers for the device data pointers. Each

[3]https://crpl.cis.udel.edu/oaccvv/

[4]https://github.com/OpenACCUserGroup/OpenACCV-V

Authorized licensed use limited to: UNIVERSITY OF DELAWARE LIBRARY. Downloaded on December 05,2023 at 00:13:08 UTC from IEEE Xplore. Restrictions apply.

of these needs to be assigned its value by passing the reference to `cudaMalloc` which allocates the data on the device. After this, we also need to copy the data to the device referenced by the device pointers. To create the kernel, we define the function that accepts references and an end bound that checks against the `blockIdx.x`, `blockDim.x`, and `threadIdx.x`. The resulting code is in Code 2. Comparing the C code and the CUDA code, the two versions of the code require totally different function calls, formatting, and syntax.

Code 3: Simple OpenACC Addition of Arrays

```
int main(){
    int N = 1<<20;
    ***Variable Declaration***
    for (int x = 0; x < N; ++x){a[x] = 10;b[x] = 15;}
    #pragma acc data copyin(a[0:N], b[0:N]) copyout(c[0:N])
        for (int x = 0; x < 1 << 14; ++x){
            #pragma acc parallel loop independent
            for (int y = 0; y < N; ++y)
                c[y] = a[y] + b[y];
}       }
```

The OpenACC version in Code 3 requires few changes to the serial code. We add a `#pragma acc data` construct, which manages the device data environment for that region, and a `#pragma acc parallel loop independent` construct. The `parallel loop` construct specifies that the following region should execute on the device and run in parallel, and the `independent` clause specifies that the loop does not have any inter-iteration dependencies.

III. Challenging Tests

In order to validate and verify the conformance of compilers to the latest OpenACC specification, the testsuite must change over time. The OpenACC V&V testsuite team is constantly updating the suite to test the most recent specification. In this section, we highlight some challenging tests within this update, specifically directives and clauses, that evoked varying interpretations.

A. Routine Directive Bind Clause and Lambda Function

The integration of additional C++ options with the `routine` directive and `bind` clause proved to be particularly tricky. These tests required several stages of development to guarantee tests exist for every combination of both C++ class objects with arrays and normal functions in tandem with lambda functions. The first obstacle is interpreting the specification. Reading over the relevant section, the `routine` directive was relatively straightforward. But for the `bind` clause, it is vague on both the implication and the use case. Thankfully, an imperfect test that utilizes these features exists but also creates a race condition. Utilizing the test to understand the directive application and implementation, the race condition was removed. This serves as a foundation for all of the other tests. Based on the specification, the `routine` directive allows the user to specify a function to be compiled and executed on both the host and on the device when the function is called on the host. The bind clause allows for further choice of which functions will be executed on the host versus the device. The next challenge was learning how to correctly implement a C++ lambda function. So, the best next step that will provide practice without the extra complexity is testing its implicit behavior within a `pragma`. Interpreting it as an alternative way to declare functions that can exist within any scope, the integration was relatively straightforward. With the implementation of the routine and the lambda function understood, this stage of testing reveals several features that are not fully implemented. The current implementation of the lambda function requires the host function to not be a lambda function, the device lambda function to be placed at least one function declaration below where the prototyped `pragma` is declared, and the accelerator routine that is a lambda function must be defined after the `pragma` is declared.

Code 4: Nonprototype Routine Declaration

```
#pragma acc routine vector bind("device_array_array")
real_t host_array_array(real_t * a, long long n){
    #pragma acc loop reduction(+:returned)
    real_t returned = 0.0;
    for (int x = 0; x < n; ++x)
        returned += a[x];
    return returned;
}
auto device_array_array = [](real_t * a, long long n){
    real_t returned = 0.0;
    #pragma acc loop reduction(-:returned)
    for (int x = 0; x < n; ++x)
        returned -= a[x];
    return returned;
};
```

Code 5: Prototype Routine Declaration

```
real_t host_array_array(real_t *a, long long n);
#pragma acc routine(host_array_array) vector bind(
     device_array_array)

real_t host_array_array(real_t * a, long long n){
    #pragma acc loop reduction(+:returned)
    real_t returned = 0.0;
    for (int x = 0; x < n; ++x)
        returned += a[x];
    return returned;
}
auto device_array_array = [](real_t * a, long long n){
    real_t returned = 0.0;
    #pragma acc loop reduction(-:returned)
    for (int x = 0; x < n; ++x)
        returned -= a[x];
    return returned;
};
```

Taking the four possible permutations, seen in Code 4 and 5, of lambda functions and a normal function, these evaluate to a total of sixteen different declarations to integrate C++ class objects. Already having experience in using class objects, the last stage of development was less challenging. The implementation revealed another error, namely, utilizing a `copy` and `copyout` clause on a class object resulted in a segmentation fault when trying to access the data. But, this was worked around by utilizing the `update` directive with the `host` clause for the purpose of this test.

B. If Clause

The `if` clause in OpenACC acts similarly to how if statements in the C language work. When the condition in the `if` clause evaluates to true, the region will execute, accelerated, on the device. When the condition evaluates to false, the region will run unaccelerated on the local host thread. Starting in the OpenACC 3.0 Specification, the `if` clause was applicable to the following directives: `init`, `set`, `shutdown`, and `wait`. The usage for the four directives is similar; the following test is for the `init` directive. When the `init` directive is called,

Authorized licensed use limited to: UNIVERSITY OF DELAWARE LIBRARY. Downloaded on December 05,2023 at 00:13:08 UTC from IEEE Xplore. Restrictions apply.

the runtime environment for a given device is initialized. When we created the test for the `init` directive with an `if` clause, we originally believed the simplest way to test this was to create a compute region and see if it ran. We created two arrays and filled the first array, A, with random values, and the other array, B, with all zeros. We looped through array A and checked if the value in A matched the same value in A, which would of course give us a true evaluation. Since the values would always match, we set the value in array B to the value in array A within an `init if pragma` that evaluated to true. We would then check if the values from both arrays always matched, and returned an error if they didn't.

Shortly after this code was created, through discussion, an issue with our test was found; the `init` directive doesn't deal with compute regions. One doesn't use `init` to compute any type of values, it is rather used to initialize the runtime for some device. Therefore, our original interpretation completely misused the directive. Connecting the `if` clause to an `init` directive only determines if runtime is initialized on either the current device or the local thread. Once we understood this, the test became much more clear, reference Code 6.

Code 6: Correct Usage of Init If

```
int device_num = acc_get_device_num(
    acc_get_device_type());
#pragma acc init if(device_num == device_num)
```

So, instead of computing values, we had to call a device type and, using the `if` clause, check if the device type we are calling matches the device type we are currently using. This is why there are two tests in this file: we had to check for a situatio-https://www.overleaf.com/project/62e30e8628736e39713d22c4n where the device types match and one where the device types do not match. The `if` clause follows a similar protocol for `set`, `shutdown`, and `wait`; a device will perform the action for the respective directive only if the current device type called matches the device type currently being used. The `if` clause is currently only implemented on the `wait` directive, but will be officially implemented on `init`, `set`, and `shutdown` in the near future.

C. Reference Counter Zero Behavior

Code 7: Reference Counter Zero

```
#pragma acc data copyin(a[0:n], b[0:n]) copy(c[0:n])
    #pragma acc parallel loop
    for (int x = 0; x < n; ++x)
        c[x] = a[x] + b[x];
#pragma acc exit data copyout(c[0:n])
for (int x = 0; x < n; ++x){
    if (fabs(c[x] - (a[x] + b[x])) > PRECISION)
        err += 1;
}
```

The goal of this test, Code 7, is to verify that an `exit data` clause or `copyout` clause does not cause a runtime error when executed on a variable whose reference count is zero. In OpenACC 3.0 and earlier versions, this action should cause a runtime error; however, in OpenACC 3.1 and later versions, the error should not occur and no action should be taken. To test this feature, a data region is created. A variable, the c array, is copied in and then out of the region. On entrance to the data region, the reference count of the variable is incremented to one, and on exit it is decremented to zero. After the data region, an `exit data copyout` directive is executed on the aforementioned variable whose reference count is now zero. This should cause no action, according to the update to the specification in OpenACC 3.1, whereas it would cause a runtime error in OpenACC 3.0 and before. This test was difficult because the reference counter is not directly accessible to the user of OpenACC; the programmer must trust the implementation of the compiler. After multiple revisions of this test, we decided the best way to test it was to use an `exit data` directive with a `copyout` clause after a data region, seen in Code 7. The redundant `copyout` clauses on the data region would not cause the reference count to go below zero. Moreover, clauses on a data region are not read by the compiler in the order that they are listed, so putting a `copyout` on a data region where the data to be copied is not already present can cause issues. The `exit data` directive moves data from device memory to host memory and deallocates the memory on the device. The data has already been copied out after the end of the data region, so the reference counter is zero. Attempting an `exit data` directive for a variable who was copied out at the end data region means that this directive is called on a variable whose reference counter is zero. This tests the specification statement that no action is taken when an `exit data` directive is performed on a variable whose reference counter is zero. In older versions of OpenACC, this action would cause an error, but should not in this update.

D. Init Directive Device Type and Num Clauses

Initially looking into the `device type` clause, the feature was relatively straightforward to utilize with the function `acc_get_device_type`. Making the assumption that the return value for the function could be used for the clause, we looked into the specification for further guidance and found that it accepts device-type-list. Looking throughout the specification, this value was never explicitly defined.

Code 8: device_type for GCC

```
int test1(){
    int err = 0;
    srand(SEED);
    int device_num = acc_get_device_num(acc_get_device_type
        ());
    #pragma acc init device_num(device_num)
    return err;}
```

But when using this implementation, Code 8, for an NVIDIA GPU, the compiler threw an error about `device type` requiring the keywords: host, multicore, default, or NVIDIA. To test this further, this code was also compiled for AMD GPUs using GCC resulting in no compilation or runtime errors. Rewriting the tests to incorporate the outlined changes from the compiler from NVIDIA, the new test, Code 9, was compiled using nvc and GCC.

Asking the OpenACC community for further guidance led to the appendix section. This section outlines the different keywords to use for the different GPU vendors. While very useful, with no mention of the multicore option for NVIDIA or the `acc_get_device_type` option for GCC, it does

Authorized licensed use limited to: UNIVERSITY OF DELAWARE LIBRARY. Downloaded on December 05,2023 at 00:13:08 UTC from IEEE Xplore. Restrictions apply.

Code 9: device_type for nvc

```
int test1(){
   int err = 0;
   srand(SEED);
   #pragma acc init device_type(host)
   #pragma acc init device_type(multicore)
   #pragma acc init device_type(default)
   #pragma acc init device_type(nvidia)
   return err;}
```

not fully encapsulate all of the possible keyword options for the GPUs. This also resulted in the encapsulation of each of the options into their own tests and the NVIDIA option being isolated into a separate test file.

IV. Infrastructure

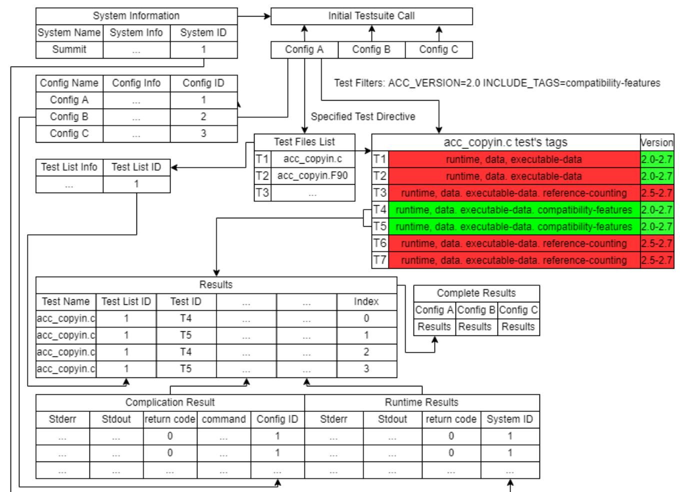
Alongside the validation suite for OpenACC, an infrastructure has been created to provide tools to maximize customizability and readability. Formatting the output and results, it was designed to handle errors at compilation and/or at runtime, simplify running the testsuite, and set up the runtime environment. Written in python using packages supported by both python 2 and python 3, it has been developed for compatibility across a majority of systems. Figure 1 gives an overview of the infrastructure.

To customize the infrastructure, the users may edit the configuration file. Within the file, documentation is provided for each feature. Some of the features are left intentionally empty to allow for the full capability of the infrastructure. The most important feature is the interchangeable compiler that is being used. For C, C++, and Fortran tests, the compiler must be specified individually. Compilation flags must also be specified as well, such as *-acc*, which is required for OpenACC libraries. The users also have the option to choose among the flags to use on the command line. These include but are not limited to: OpenACC-specific flags like *-Minfo=all*, displaying error messages during runtime or compilation, and output of results for each test file. It is also possible to specify commands to be run pre- or post compilation or execution if the user would like to further customize the environment or output format.

Next, the users can partition the testsuite to only execute a certain subsection of tests. This could be accomplished in one of two ways: running only a specified directory, or using the tag-based system for conditional compilation. The directory method utilizes the organization of the testsuite to isolate a specific version of OpenACC tests to execute. For the tag-based system, each test is tagged at the beginning of the code based on what feature of OpenACC it is testing. It can also be used to isolate a version of OpenACC that the tests are valid in. If tests are tagged with a later version, they will be omitted. For both methods, the directory where the tests are built can be specified. When no specified directory or tag is provided, the infrastructure compiles and executes all of the tests within the validation suite.

Another important feature of the configuration file is the ability to customize how the output is formatted by utilizing one of the current options: json, txt, or html. With the json option, the data is exported in json format with nothing

Fig. 1: Overview of the infrastructure



omitted. With the txt option, a list of commands and results are output within a text log. The html option will produce an altered json format for the purpose of being used in an html page. This altered json format can be read with the use of the results template.

Lastly, the testsuite could take several hours to fully execute all of the tests provided. To cut down on the execution time, a timeout parameter feature is provided to prevent the suite from hanging in execution when a test hangs. Each test should complete within 5 seconds, so a timeout of 10 seconds is recommended. However, the largest time complexity is n^3 for any test currently, so the timeout can be scaled proportionally if the problem size is changed by the user.

Table 1: Compilation Flags

Compilation Flags	description
-c	specify one or more configuration files
-env	environmental output
-o	naming the output file
-verbose	interactive intuitive all
-system	label system being used

Table 1 describes some of the flags available to users. A comprehensive list of all the commands and how to run the testsuite can be found on the OpenACC V&V Testsuite GitHub [5].

V. LLVM Integration

The primary goal of the integration is to facilitate the use of the OpenACC V&V Suite for testing the behavior of Clacc's current and future OpenACC support [19].

We anticipate this testsuite project will immediately benefit the development of Clacc (OpenACC support for C/C++ in Clang) and Flacc (OpenACC support for Fortran in Flang).

[5]https://github.com/OpenACCUserGroup/OpenACCV-V

Authorized licensed use limited to: UNIVERSITY OF DELAWARE LIBRARY. Downloaded on December 05,2023 at 00:13:08 UTC from IEEE Xplore. Restrictions apply.

These efforts would also be beneficial to establish feedback from Clacc and Flacc to the OpenACC V&V testsuite team to advance testing for the latest OpenACC specification versions implemented in LLVM. Clacc and Flacc are developing their own test suites that are to be contributed to upstream LLVM. While the OpenACC V&V Suite can be used for validating all OpenACC implementations, the Clacc and Flacc testsuites are not meant to be that general. For example, Clacc's test suites exercise Clacc-specific command-line options and source-to-source capabilities, which are beyond the scope of the OpenACC specification. The major value of the OpenACC V&V Suite would be an objective, third-party assessment of LLVM's (Clacc) conformance to the OpenACC specification that can be compared with other OpenACC implementations. LLVM has a testing infrastructure that contains regression tests and a collection of whole programs [20], all of which are driven by the LLVM Integrated Tester (LIT) tool. Clacc's and Flacc's own test suites are being developed as extensions of LLVM's regression tests, which are small tests that are expected to always pass and should be run before every commit. The OpenACC V&V suite would be integrated into the collection of whole programs.

A. Implementation

To set up an environment for the OpenACC validation and verification suite entails a set of requirements that justify the different features that we have implemented so far. These requirements are as follows:

1) As mentioned in [21] the Clacc compiler translates OpenACC to OpenMP in order to build upon the OpenMP support being developed for Clang and LLVM. To maximize reuse of the OpenMP implementation, Clacc performs this translation early, on the abstract syntax tree, which is the compiler front end's internal representation of the source code. This translation is effectively a lowering of the representation and thus follows the traditional ordering of compiler phases.
2) A CMakeList file has been created and included in this project. It is a generator of build systems and used as an entry point to our testsuite. It generates the Makefile which allows users to compile, run, and report test results. A set of make rules has been created for each purpose, together with a set of options that modify each rule's behavior.
3) Those who would like to use this testsuite must be able to obtain and export compilation and results. Hence, the designed infrastructure should allow them to obtain results in either a json format, or format for exporting to other analysis tools and scripts.

Clacc is a project to develop OpenACC compiler and runtime support for C and C++ in Clang and LLVM. Clacc's source code is available publicly on GitHub as part of the LLVM DOE Fork maintained by ORNL. Clacc should be built in the same manner as upstream LLVM when Clang and OpenMP support are desired [21]. This integration includes a CMakeLists for the OpenACC V&V suite so it can be built as part of the LLVM test-suite. The CMakeLists will search for all C and C++ source files of the OpenACC V&V suite, and compile and run them. Running llvm-lit (or "make check") will require a compatible accelerator on the running machine. LLVM OpenACC V&V Integration can be found [6]. With these requirements defined, we present our infrastructure in the rest of this section.

VI. Setup and Compilation Flags

Table 2: Compiler Versions and Platforms

System	Hardware	Compiler	Flags
DARWIN	NVIDIA T4	nvc 21.9	-acc=gpu
DARWIN	NVIDIA T4	nvc 22.5	-acc=gpu
DARWIN	NVIDIA Tesla T4	GCC 10.1.0	-fopenacc
DARWIN	NVIDIA Tesla T4	GCC 11.3.0	-fopenacc
DARWIN	NVIDIA Tesla T4	GCC 12.1.0	-fopenacc
Summit	NVIDIA Tesla V100	nvc 20.9	-acc=gpu
Summit	NVIDIA Tesla V100	nvc 22.5	-acc=gpu
Summit	NVIDIA Tesla V100	GCC 10.2.0	-fopenacc
Summit	NVIDIA Tesla V100	GCC 12.1.0	-fopenacc
Perlmutter	NVIDIA A100 Tensor Core	nvc 21.11	-acc=gpu
Perlmutter	NVIDIA A100 Tensor Core	nvc 22.5	-acc=gpu
Perlmutter	NVIDIA A100 Tensor Core	GCC 10.3.0	-fopenacc
Perlmutter	NVIDIA A100 Tensor Core	GCC 11.2.0	-fopenacc
Perlmutter	NVIDIA A100 Tensor Core	Clacc Git # "4879e9"	-fopenacc
Spock	AMD MI100	GCC 10.3.0	-fopenacc
Spock	AMD MI100	GCC 11.2.0	-fopenacc
Spock	AMD MI100	HPE 12.0.0	-hacc,noomp
Spock	AMD MI100	HPE 13.0.0	-hacc,noomp

Table 2 displays the various systems and compilers we used, as well as the hardware associated with each system and what flags were used to specify OpenACC. You can find more information on the system on our website. For more information on command-line options for running OpenACC applications and installation of Clacc, please go to the Clacc Github repository of the llvm-project.[7]

[6]https://github.com/llvm/llvm-test-suite/tree/main/External
[7]https://github.com/llvm-doe-org/llvm-project/tree/clacc/main

Authorized licensed use limited to: UNIVERSITY OF DELAWARE LIBRARY. Downloaded on December 05,2023 at 00:13:08 UTC from IEEE Xplore. Restrictions apply.

Fig. 2: Verification and Validation Testsuite Results in C

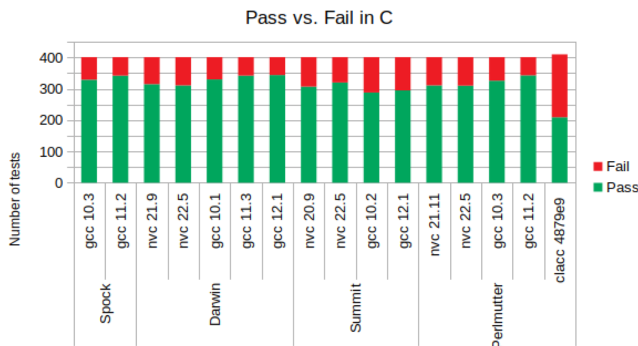
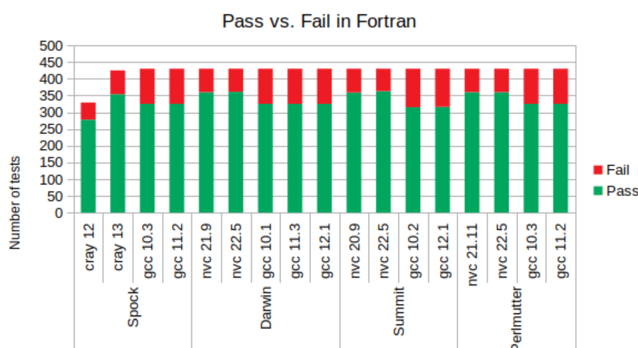


Fig. 3: Verification and Validation Testsuite Results in Fortran



VII. Results

(Note: The compiler and their versions, the number of tests, and the OpenACC specification version used to generate results for this paper are as current as Sept 2022 (at the time of submitting this paper).)

While hundreds of tests have been added to or modified in the testsuite, nothing has changed with the portability and ease of use. After adjusting to each system the suite has run with multiple compilers on multiple systems, when compatible, as described in the previous section: Setup and Compilation Flags in Section VI. Over the past couple of years, the testsuite has grown to a suite comprised of more than 800 tests. Running these tests across multiple systems gave varying results. Out of the 831 tests, approximately 80% passed with the NVIDIA and GNU compilers. These results are displayed in Figures 2 and 3. Due to the renewed HPE Cray OpenACC compiler now focusing on the Fortran language, the success rate is higher at roughly 84%.

The success rate of each compiler will continue to evolve as more features are implemented per compiler and more bugs are fixed. We hope that the testsuite not only provides a metric for compiler teams to measure their current implementation against, but also will be a tool in the future for compilers to verify that features continue to work as changes continue to be implemented.

A. NVIDIA Compiler's Conformance to the OpenACC Specifications

The NVIDIA compiler was tested on three different systems intentionally to identify variations in results. Utilizing the latest available compiler version on each system, namely 22.5, and an older version to demonstrate validity over time, no significant variation between versions tested was seen. On average in these results, the NVIDIA compiler supports 80.7% of features currently implemented into the testsuite which includes most features up to OpenACC 3.2.

Some of the failures are due to issues in the implementations of OpenACC; issues have been reported to NVIDIA who have logged bugs. These include but are not limited to: the `init` directive with the `device_type` clause, a combined construct for `serial loop` with the `vector`, `worker`, and `gang` clauses. Other failures are because some features past OpenACC 2.7 are not yet implemented. The `if` clause on the `init`, `set`, and `shutdown` directives to name a few. NVIDIA is aware of all failures and are continually working to improve the compiler.

B. The GNU Compiler's Conformance to the OpenACC Specifications

The GNU compilers support anywhere between 78.3% and 80.3% of the features currently implemented. The only outlier was on Summit, which supported between 72.6% and 73.5% of features currently implemented with GNU compilers. These features in the testsuite include most of the features implemented up to OpenACC 3.2 and were written in C and Fortran. While the number of tests written in C that passed varied per machine, the results were consistent with each other. The Fortran tests had almost the exact same pass rate, with again Summit being the only exception for files in both languages. Most of these failures were due to either bugs in the compiler or the compiler not fully implementing all features up to OpenACC 3.2.

C. The HPE Cray Fortran Compiler's Conformance to the OpenACC Specifications

One of the recent additions to the suite is HPE Cray's OpenACC support for Fortran. To note, this compiler focuses on the Fortran language, thus limiting the tests we were able to run. Our infrastructure will exclude the other two languages when running, the option is within the config file. Once excluded, the testsuite is left with a total number of 329 tests in Fortran with a success rate of approximately 84%. To achieve this success rate we used the pre-exascale system called Spock [22] at Oak Ridge National Laboratory. This system has nodes with AMD GPUs MI60 and MI100. Our tests ran on the MI100.

The HPE Cray compiler uses PTX (Parallel Thread Execution) instructions which are translated into assembly, it does not translate into CUDA. However, as clarified in previous sections, the Cray compiler fully supports OpenACC 2.0, with partial support for OpenACC 2.6. This includes many behaviors like `copy` data clauses and the

Authorized licensed use limited to: UNIVERSITY OF DELAWARE LIBRARY. Downloaded on December 05,2023 at 00:13:08 UTC from IEEE Xplore. Restrictions apply.

`default` clause. Our testsuite has confirmed support for the `atomic` construct, the `reduction` clause with the `kernels` and `parallel` constructs. Notably, since partial support is up to 2.6 some test cases will fail. Some of these cases include: `acc_on_device`, `acc_set_device_num`, and `acc_set_device_type`. We cannot guarantee that these tests will succeed in the next compiler update. We can guarantee that our team will continue to look in detail and communicate with the developer about these results.

D. The LLVM/Clacc Compiler's Conformance to the OpenACC Specifications

The Clacc compiler focuses on the C and C++ languages, resulting in a limited number of tests able to run. Our infrastructure will exclude the Fortran language when running; the option is within the config file. Clacc was tested on NERSC's Perlmutter using the LLVM GitHub commit hash "4879e96", which passed approximately 50% of OpenACC V&V suite's tests. However, 99.52% tests passed out of all features implemented by Clacc. These features in the testsuite include most of the features implemented up to OpenACC 3.2 and were written in C. Only one test failure was due to a bug in Clacc. All remaining failures were due to OpenACC features yet to be implemented by the compiler, including `kernels`, `serial`, and `async/wait`. The Clacc compiler has greatly increased their support for OpenACC lately, but this is a work in progress.

VIII. Discussion

This project has demonstrated the conformance of different versions of OpenACC compilers and their current performance targeting different systems including OLCF's Summit and Spock, NERSC's Perlmutter, and the University of Delaware's Darwin. The previously published version of this effort [23] had deemed the HPE Cray and Clacc compilers were not ready for validation. However, the two compilers are now showing increased support for OpenACC.

We were able to target multiple systems due to the infrastructure that we have built since our last publication. Because the number of tests within our suite has increased and many users may want to run the full suite, the infrastructure will be the easiest way to validate any compiler on any system. One thing to note, we want to update our implementation of the infrastructure to better run C++ tests as currently, the infrastructure leaves these tests out of the results.

Recently, features newly added in versions beyond 3.0 may not have compiler implementations as of yet. However, we have written the tests awaiting implementation. For example, the `if` clause was recently added to the `init`, `shutdown`, `set`, and `wait` directives in OpenACC 3.0. At the moment of this publication, only the wait directive with the `if` clause has been implemented by the NVIDIA (nvc) compiler. `Init`, `shutdown`, and `set` have yet to be implemented but the testsuite contains a test for each of these three combinations. This is the major reason why new tests are failing; however, as compilers will develop they will start to implement these newer features. Thus, our test will validate the compiler implementation of those newly compiler-implemented features.

IX. Related Work

In our past publications on this topic, we have written about the testsuite covering OpenACC 1.0 [18] and OpenACC 2.5 [23]. These two publications are the most related to our ongoing work. There is testsuite similar to ours, at the University of Delaware, which targets the OpenMP programming model. Their validation and verification testsuite [24] also aims to check for conformance of features in compiler implementations. Since 2003, the OpenMP testsuite team has developed tests according to OpenMP 2.0 [25], version 2.5 [26], version 3.1 [27] and version 4.5 [28], [29]. The team works on the next publication as we speak.

Other efforts towards creation of a validation and verification testsuite include Csmith [30] and the Parallel Loops testsuite [31], modeled after the Livermore Fortran kernels [32] as also mentioned in our previous publication. Csmith uses differential testing to perform a randomized test-case generator exposing compiler bugs and the Parallel Loops testsuite chooses a set of routines to test the strength of a computer system (compiler, run-time system, and hardware) in a variety of disciplines. OvO [33] is another suite that is a collection of OpenMP offloading tests for C++ and FORTRAN. OvO is focused on testing extensively hierarchical parallelism and mathematical functions.

X. Conclusion and Future Work

The purpose of this project is to develop test cases in order to validate and verify compilers' implementations of OpenACC features. Compilers and the testsuite must advance alongside the evolving OpenACC specification. The evolution of the testsuite gives the opportunity to create test cases and corner cases resulting in the finding of bugs, thus, improving the compilers' implementation of each feature.

In the future, one goal is to create an example guide for the OpenACC community. Taking inspiration from OpenMP 4.5 Examples, our hope is that the guide will display relatively simple examples of features in OpenACC. We will kick off this process by first surveying the OpenACC user community and seeking their input on features they would like us to create examples for. The goal of this document would be to explain each directive and clause with a short statement outlining what the code is accomplishing. Automating the process gives the opportunity to develop the guide at a faster pace; however, our main focus, first, is to present a guide of the utmost quality to the community. Our main hope for the example guide is to provide sample code for OpenACC users to get a basic understanding of applying OpenACC to their codes.

In order to ensure that compiler implementations are as bug-free as possible and conform to the OpenACC specification, we will add more rigorous testing to cover edge cases that may not be possible with basic tests. To make the testsuite easy to use, we will create a transposable connection between the testsuite and the specification. The testsuite repository will

Authorized licensed use limited to: UNIVERSITY OF DELAWARE LIBRARY. Downloaded on December 05,2023 at 00:13:08 UTC from IEEE Xplore. Restrictions apply.

have tests that are related to the specification, and each test will be tagged with the corresponding definition.

The results shown have been taken relatively close to the submission date of this paper. We will continue to develop the testsuite with valuable feedback and input from our mentors. For more details on up-to-date results with new compilers, compiler versions, up-to-date tests or to contribute to this project please consult our website[8].

A. Acknowledgments

We are grateful to OpenACC for their support on this project, including technical support from Mathew Colgrove, Jeff Larkin, Duncan Poole, Christophe Harle, Guray Ozen, Wael Elwasif, Seyong Lee and Joel Denny for continued help with this project.

This research was supported by the National Science Foundation (NSF) under grant no. 1919839, in part through the use of DARWIN computing system: DARWIN – A Resource for Computational and Data-intensive Research at the University of Delaware. This material is also based upon work supported by NSF under grant no. 1814609.

This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231 using NERSC award.

References

[1] OpenACC, "OpenACC, Directives for Accelerators," http://www.openacc.org/.

[2] OpenMP, "Openmp 4.5 specification."

[3] NVIDIA, "CUDA SDK Code Samples," http://developer.nvidia.com/cuda-cc-sdk-code-samples, accessed: 2017-02-03.

[4] OpenCL, "OpenCL," https://www.khronos.org/.

[5] "NVIDIA Thrust," https://developer.nvidia.com/thrust, accessed: 2017-02-03.

[6] H. C. Edwards, C. R. Trott, and D. Sunderland, "Kokkos: Enabling manycore performance portability through polymorphic memory access patterns," *Journal of Parallel and Distributed Computing*, vol. 74, no. 12, pp. 3202–3216, 2014.

[7] S. Lee, J. Kim, and J. S. Vetter, "OpenACC to fpga: A framework for directive-based high-performance reconfigurable computing," in *Parallel and Distributed Processing Symposium, 2016 IEEE International*. IEEE, 2016, pp. 544–554.

[8] "Gcc supporting openacc model," https://gcc.gnu.org/wiki/OpenACC.

[9] J. E. Denny, S. Lee, and J. S. Vetter, "Clacc: Translating openacc to openmp in clang," in *2018 IEEE/ACM 5th Workshop on the LLVM Compiler Infrastructure in HPC (LLVM-HPC)*. IEEE, 2018, pp. 18–29.

[10] A. Tabuchi, M. Nakao, and M. Sato, "A source-to-source openacc compiler for cuda," in *European Conference on Parallel Processing*. Springer, 2013, pp. 178–187.

[11] S. Lee and J. S. Vetter, "Openarc: Open accelerator research compiler for directive-based, efficient heterogeneous computing," in *Proceedings of the 23rd international symposium on High-performance parallel and distributed computing*, 2014, pp. 115–120.

[8]https://crpl.cis.udel.edu/oaccvv

[12] M. Wolfe, S. Lee, J. Kim, X. Tian, R. Xu, S. Chandrasekaran, and B. Chapman, "Implementing the openacc data model," in *2017 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*. IEEE, 2017, pp. 662–672.

[13] S. Sathe, "Accelerating the ansys fluent r18. 0 radiation solver with openacc," 2016.

[14] Quantum Chemistry (QC) on GPUs, https://images.nvidia.com/content/tesla/pdf/quantum-chemistry-may-2016-mb-slides.pdf.

[15] W. Sawyer, G. Zaengl, and L. Linardakis, "Towards a multi-node openacc implementation of the icon model," in *EGU General Assembly Conference Abstracts*, 2014, p. 15276.

[16] X. Lapillonne and O. Fuhrer, "Using compiler directives to port large scientific applications to gpus: An example from atmospheric science," *Parallel Processing Letters*, vol. 24, no. 01, p. 1450003, 2014.

[17] J. Y. Kim, J.-S. Kang, and M. Joh, "Gpu acceleration of mpas microphysics wsm6 using openacc directives: Performance and verification," *Computers & Geosciences*, vol. 146, p. 104627, 2021.

[18] C. Wang, R. Xu, S. Chandrasekaran, B. Chapman, and O. Hernandez, "A validation testsuite for OpenACC 1.0," in *Parallel & Distributed Processing Symposium Workshops (IPDPSW), 2014 IEEE International*. IEEE, 2014, pp. 1407–1416.

[19] "OpenACC VV suite for llvm," https://docs.google.com/document/d/1s2txylinBjCHYXLJcUdqM_0XvT7qk-ajJCfMQdqF9jo/edit#heading=h.gg7fwklw6qfs.

[20] "LLVM testing infrastructure guide¶," https://llvm.org/docs/TestingGuide.html.

[21] C. Coti, J. E. Denny, K. Huck, S. Lee, A. D. Malony, S. Shende, and J. S. Vetter, "Openacc profiling support for clang and llvm using clacc and tau," in *2020 IEEE/ACM International Workshop on HPC User Support Tools (HUST) and Workshop on Programming and Performance Visualization Tools (ProTools)*. IEEE, 2020, pp. 38–48.

[22] "Spock quick-start guide," https://docs.olcf.ornl.gov/systems/spock_quick_start_guide.html.

[23] K. Friedline, S. Chandrasekaran, M. G. Lopez, and O. Hernandez, "Openacc 2.5 validation testsuite targeting multiple architectures," in *High Performance Computing*, J. M. Kunkel, R. Yokota, M. Taufer, and J. Shalf, Eds. Cham: Springer International Publishing, 2017, pp. 557–575.

[24] OpenMP Validation and Verification Suite, "Openmp 3.1 Specification," https://github.com/sunitachandra/omp-validation.

[25] M. Müller and P. Neytchev, "An openmp validation suite," in *Fifth European Workshop on OpenMP, Aachen University, Germany*, 2003.

[26] M. Müller, C. Niethammer, B. Chapman, Y. Wen, and Z. Liu, "Validating openmp 2.5 for fortran and c/c," in *in Sixth European Workshop on OpenMP, KTH Royal Institute of Technology*. Citeseer, 2004.

[27] C. Wang, S. Chandrasekaran, and B. Chapman, "An openmp 3.1 validation testsuite," in *International Workshop on OpenMP*. Springer, 2012, pp. 237–249.

[28] J. M. Diaz, S. Pophale, O. Hernandez, D. E. Bernholdt, and S. Chandrasekaran, "Openmp 4.5 validation and verification suite for device offload," in *International Workshop on OpenMP*. Springer, 2018, pp. 82–95.

[29] J. M. Diaz, K. Friedline, S. Pophale, O. Hernandez, D. E. Bernholdt, and S. Chandrasekaran, "Analysis of openmp 4.5 offloading in implementations: correctness and overhead," *Parallel Computing*, vol. 89, p. 102546, 2019.

[30] X. Yang, Y. Chen, E. Eide, and J. Regehr, "Finding and understanding bugs in c compilers," in *ACM SIGPLAN Notices*, vol. 46, no. 6. ACM, 2011, pp. 283–294.

[31] J. Dongarra, M. Furtney, S. Reinhardt, and J. Russell, "Parallel loops—a test suite for parallelizing compilers: Description and example results," *Parallel Computing*, vol. 17, no. 10-11, pp. 1247–1255, 1991.

[32] F. H. McMahon, "The livermore fortran kernels: A computer test of the numerical performance range," Lawrence Livermore National Lab., CA (USA), Tech. Rep., 1986.

[33] O. S. T. a Subset of OpenMP Offload, https://www.openmp.org/wp-content/uploads/Applencourt_OvO_slide.pdf.

Authorized licensed use limited to: UNIVERSITY OF DELAWARE LIBRARY. Downloaded on December 05,2023 at 00:13:08 UTC from IEEE Xplore. Restrictions apply.

APPENDIX

A. Summarize the experiments reported in the paper and how they were run.

We ran a suite of hundreds of tests to measure conformance of compilers to OpenACC on a variety of hardware: Summit: (2) 22 core IBM POWER9 processors CPUs, (6) NVIDIA Tesla V100 GPUs Spock: (1) 64-core AMD EPYC 7662 "Rome" CPU, (4) AMD MI100 GPUs Perlmutter: (1) AMD Milan CPU, (4) NVIDIA A100 Tensor Core GPUs Darwin: (2) 32-core AMD EpycTM 7502 processors, (1) NVIDIA T4 GPU The testsuite was run using an infrastructure script to format results and provide portability, reproducibility, and simplicity. The testsuite and infrastructure used to run it and format results are freely available (Github link below). The tests are based off of the most recent version the OpenACC specification at this time, version 3.2, which is available to the public (linked below). For the various targets, we used: NVHPC 20.9, 21.9, 21.11, and 22.5; GCC 10.1, 10.2, 10.3, 11.2, 11.3, and 12.1; and HPE Cray 12.0 and 13.0.

B. Software Artifact Availability:

All author-created software artifacts are maintained in a public repository under an OSI-approved license.

C. Data Artifact Availability:

All author-created data artifacts are maintained in a public repository under an OSI-approved license.

D. URL/DOI List (separate line per URL/DOI)

https://zenodo.org/record/7200141
https://crpl.cis.udel.edu/oaccvv/

E. Relevant hardware details:

Summit: (2) 22 core IBM POWER9 processors CPUs, (6) NVIDIA Tesla V100 GPUs Spock: (1) 64-core AMD EPYC 7662 "Rome" CPU, (4) AMD MI100 GPUs Perlmutter: (1) AMD Milan CPU, (4) NVIDIA A100 Tensor Core GPUs Darwin: (2) 32-core AMD EpycTM 7502 processors, (1) NVIDIA T4 GPU

F. Operating systems and versions:

Summit: Red Hat Fedora version 8.2 with kernel version 4.18.0. Spock: OpenSUSE version 15.2 with kernel version 5.3.18. Perlmutter: OpenSUSE version 15.3 with kernel version 5.3.18. Darwin: CentOS 7 with kernel version 3.10.0.

G. Compilers and versions:

(GNU gcc 10.1, 10.2, 10.3, 11.2, 11.3, 12.1) (NVC 20.9, 21.9, 21.11, 22.5) (HPE Cray 12, 13) Clang clacc GitHub #4879e96.

H. Libraries and versions:

(CUDA Versions: 11.0,11.3,11,7) Clacc's version of LLVM's OpenMP runtime, OpenACC Runtime Library API

I. Performed verification and validation studies:

The testsuite was run using various compiler versions on four architectures to validate compiler conformance to OpenACC specification across multiple systems and versions.

J. Validated the accuracy and precision of timings:

Each test was run using various compiler versions on four architectures to validate compiler conformance to OpenACC specification across multiple systems and versions.

K. Quantified the sensitivity of your results to initial conditions and/or parameters of the computational environment:

The testsuite was run using various compiler versions on four architectures to validate compiler conformance to OpenACC specification across multiple systems and versions.

Authorized licensed use limited to: UNIVERSITY OF DELAWARE LIBRARY. Downloaded on December 05,2023 at 00:13:08 UTC from IEEE Xplore. Restrictions apply.