




BTOR2MLIR: A Format and Toolchain for Hardware Verification

Joseph Tafese 
 University of Waterloo
 Waterloo, Canada
 jetafese@uwaterloo.ca

Isabel Garcia-Contreras 
 University of Waterloo
 Waterloo, Canada
 igarcia@uwaterloo.ca

Arie Gurfinkel 
 University of Waterloo
 Waterloo, Canada
 agurfink@uwaterloo.ca

Abstract—Formats for representing and manipulating verification problems are extremely important for supporting the ecosystem of tools, developers, and practitioners. A good format allows representing many different types of problems, has a strong toolchain for manipulating and translating problems, and can grow with the community. In the world of hardware verification, and, specifically, the Hardware Model Checking Competition (HWMCC), the BTOR2 format has emerged as the dominating format. It is supported by BTOR2TOOLS, verification tools, and Verilog design tools like Yosys. In this paper, we present an alternative format and toolchain, called BTOR2MLIR, based on the recent MLIR framework. The advantage of BTOR2MLIR is in reusing existing components from a mature compiler infrastructure, including parsers, text and binary formats, converters to a variety of intermediate representations, and executable semantics of LLVM. We hope that the format and our tooling will lead to rapid prototyping of verification and related tools for hardware verification.

I. INTRODUCTION

Hardware Verification has been one of the biggest drivers of formal verification research [1], with a history that spans many breakthroughs. The developments in this field have thrived through organized events such as the Hardware Model Checking Competition (HWMCC) [2] which has run since 2011. BTOR2 [3] has emerged as the dominating format in this competition. BTOR2 has been translated into several languages, for example, Constrained Horn Clauses (CHCs)^{1,2} and LLVM-IR³ to make use of existing verification techniques. Universality, however, was not an objective of these projects, and thus, for these translations, be it to CHCs or to LLVM-IR, similar tasks had to be replicated.

During the past decade, the LLVM project [4] has dedicated significant effort to universality. One such effort is MLIR [5], a project that proposes a generic intermediate representation with operations and types common to many programming languages. MLIR was designed to be easily extensible, by providing tools to build new intermediate representations (IR) as dialects of the base MLIR. This eases the creation of new compilers, circumventing the need to re-implement core technologies and optimizations. Extensibility and scalability are what MLIR strives for, making it a great candidate for the

creation of new tools and formats that represent many types of problems and have strong tool support for manipulating and translating problems.

During the same time, with the rise of LLVM as a compiler infrastructure, many software verification tools have been built for LLVM-IR programs. Existing tools tackle this hard problem in many ways. For example, dynamic verification is implemented in LIBFUZZER [6], a fuzzer, and KLEE [7], a symbolic execution engine; SMT-based static verification is implemented in SEAHORN [8] both as Bounded and Unbounded Model Checking; and CLAM [9] static analysis that analyzes LLVM-IR statically using abstract interpretation.

This paper contributes BTOR2MLIR, a format and toolchain for hardware verification. It is built on MLIR to incorporate advances and best practices in compiler infrastructure, compiler design, and the maturity of LLVM. At its core, BTOR2MLIR provides an intermediate representation for BTOR2 as an MLIR dialect. This dialect has an encoding very close to BTOR2 and preserves BTOR2's semantics. This design not only facilitates the creation of a new format for hardware verification but also simplifies the extension of this format to support future targets by using MLIR for all intermediate representations. For example, BTOR2MLIR can be used to generate LLVM-IR from our custom MLIR dialect. The value of this approach is quite evident in CIRCT [10], an open-source project, that applies this design to tackle the inconsistency and usability concerns that plague tools in the Electronic Design Automation industry. Although it has a different goal than BTOR2MLIR, both projects draw great benefit from adapting the benefits of an MLIR design to their respective fields.

As an added bonus, using BTOR2MLIR to generate LLVM-IR enables the reuse of established tools to apply software verification techniques to verify hardware circuits. To illustrate the usability of the toolchain, a new model checker is developed using SEAHORN. The results are compared to BTORMC [3], a hardware model checker provided by the creators of BTOR2.

The rest of the paper is organized as follows. Section II lays some background. Our format and toolchain, BTOR2MLIR, is described in Section III. We discuss its correctness in Section IV and evaluate the tool in Section V. We close with a note on related works in Section VI and conclude in

¹<https://github.com/zhanghongce/HWMCC19-in-CHC>

²<https://github.com/stepwise-alan/btor2chc>

³<https://github.com/stepwise-alan/btor2llvm>

II. BACKGROUND

BTOR2: BTOR2 [3] is a format for quantifier-free formulas and circuits over bitvectors and arrays, with SMT-LIB [11] semantics, that is used for hardware verification. BTOR2 files are often generated using tools like YOSYS [12], from the original design in a language like VERILOG [13]. A simple 4-bit counter is shown in Fig. 1. Its corresponding description, in VERILOG, is shown in Fig. 1b. The circuit updates its output at each step starting from 0 to its maximum value, 15. It also has the safety property that the output should not be equal to 15, shown by the assertion in Fig. 1b. The circuit together with the desired safety property are captured in BTOR2 in Fig. 1c. First, a bitvector of width 4 is defined as '1' in line 1. Sorts are used later when declaring registers and operations. For example, lines 2, 3, 5, and 8 refer to sort '1', respectively, by declaring '2' to be a zero bitvector (0000) (line 2), state `out` to be a register of sort '1' (line 3), '5' to be a one bitvector (0001) (line 5) and '8' to be bitvector of ones (1111) (line 8). On line 4, `out` is initialized with value '2'. On line 7, the transition function is defined (activated at each clock edge), by assigning the next state of `out` to the value `out` incremented by one (the result of line 6). Finally, a safety property is defined in line 11 with the keyword `bad`, requiring that the equality of line 10 does not hold. That is, the value of `out` is never 1111. Note that no clock is specified in Fig. 1c. In BTOR2 it is always assumed that there is one single clock, and the keyword `next` is used to declare how registers are updated after a clock cycle. For a register that has not been assigned a next value, it will get a new non-deterministic value or keep its initial value (if one was given).

BTORMC: BTORMC [3] is a bounded model checker (BMC) for BTOR2. BTORMC generates verification conditions as SMT formulas and uses BOOLECTOR [3] as an SMT solver. Based on the satisfiability result of the formula, BTORMC on our example tells us that the safety property is violated, as expected, since `out` does reach a state with value 1111.

MLIR: Multi-Level Intermediate Representation (MLIR) [5] is a project that was developed for TensorFlow [14] to address challenges faced by the compiler industry at large: modern languages end up creating their own high-level intermediate representation (IR) and the corresponding technologies. Furthermore, these domain-specific compilers have to be recreated for different compilation and optimization targets and do not easily share a common infrastructure or intermediate representations. To remedy this, MLIR facilitates the design and implementation of code generators, translators, and optimizers at different levels of abstraction and also across application domains, hardware targets, and execution environments.

Modern languages vary in the set of operations and types that they use, hence the need to create domain-specific high-level IRs. MLIR addresses this problem by making it easy for a user to define their own dialects. An MLIR dialect captures

the operations and types of a target language. It is created using TABLEGEN, a domain specific language for defining MLIR dialects. It is used to automatically generate code to manipulate the newly defined dialect including its Abstract Syntax Tree (AST) and parsing. MLIR tools and optimizations such as static single assignment, constant propagation, and dead-code elimination can be applied off the shelf to custom MLIR dialects. These capabilities make MLIR a reusable and extensible compiler infrastructure. One of its strengths is the builtin dialects it introduces, such as a BUILTIN, STANDARD, and LLVM dialects⁴, among others. These dialects make it possible to have a rich infrastructure for dialect conversion that enables a user to define pattern-based rewrites of operations from one dialect to another. For example, a dialect conversion pass is provided to convert operations in the STANDARD dialect to operations in the LLVM dialect. MLIR also provides an infrastructure for user-defined language translation passes. One such pass that is provided out of the box is a translation from LLVM dialect to LLVM-IR.

III. BTOR2MLIR

We present our tool, BTOR2MLIR, which contributes the BTOR DIALECT, and three modules on the existing MLIR infrastructure: a BTOR2 to BTOR DIALECT translation pass, a BTOR DIALECT to BTOR2 translation pass and a dialect conversion pass from BTOR DIALECT to LLVM dialect. Our tool has approximately 3900 lines of C++ code and 1200 lines of TABLEGEN. Fig. 2 shows the architecture of our tools with our contributions highlighted in green. BTOR2MLIR uses the original BTOR2 parser provided in BTOR2TOOLS [3], marked in blue, and MLIR builtin passes, marked in brown. BTOR2MLIR is open-sourced and publicly available on GitHub⁵.

We illustrate how each of the components of BTOR2MLIR works by translating a factorial circuit, shown in Fig. 3a, that is described in BTOR2. There are two safety properties, one per `bad` statement. Line 14 states that the loop counter, `i`, reaches 15. Line 19 states that the value of `factorial` is always even.

BTOR DIALECT: Our first contribution is the BTOR DIALECT, an MLIR dialect to represent BTOR2 circuits. Fig. 3b shows the BTOR DIALECT code corresponding to Fig. 3a. It represents the execution of the circuit using an MLIR function `main`. The control flow is explicit, using a standard MLIR representation of basic blocks with arguments and branches. The example has two basic blocks: an unnamed initial block (`bb0`) and a block `bb1`. Circuit initialization is modeled by instructions in `bb0`, and each cycle by instructions in `bb1`. Note that `bb1` has two predecessors: `bb0` for initialization and `bb1` for each cycle. Bitvector types are mapped to integer types (provided by MLIR), for example, `bitvec 4` becomes `i4`. Each operation in the BTOR DIALECT, prefixed with `btor`, models a specific BTOR2 operation. For example, `btor.mul`

⁴<https://github.com/llvm/llvm-project/tree/release/14.x/mlir/include/mlir/Dialect>

⁵<https://github.com/jetafese/btor2mlir/tree/llvm-14>

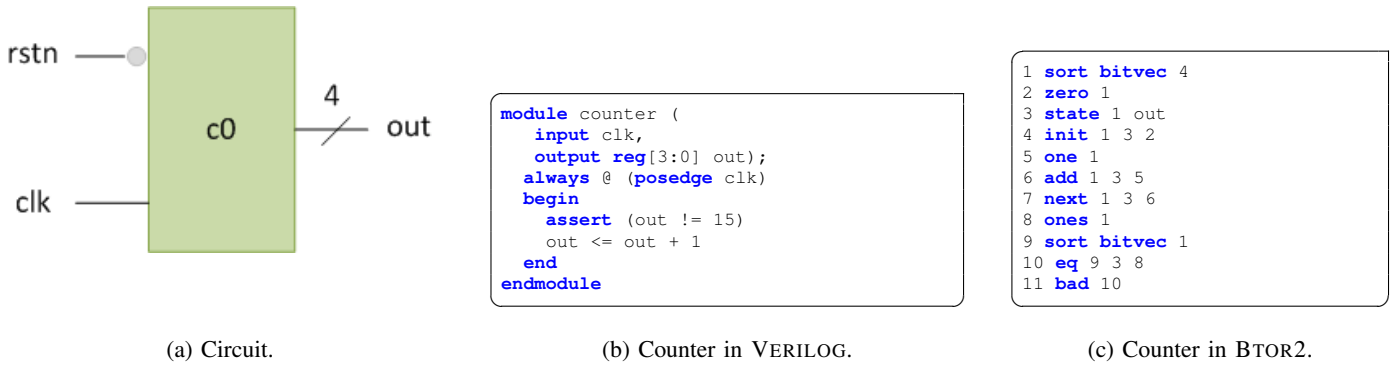


Fig. 1: 4-bit counter.

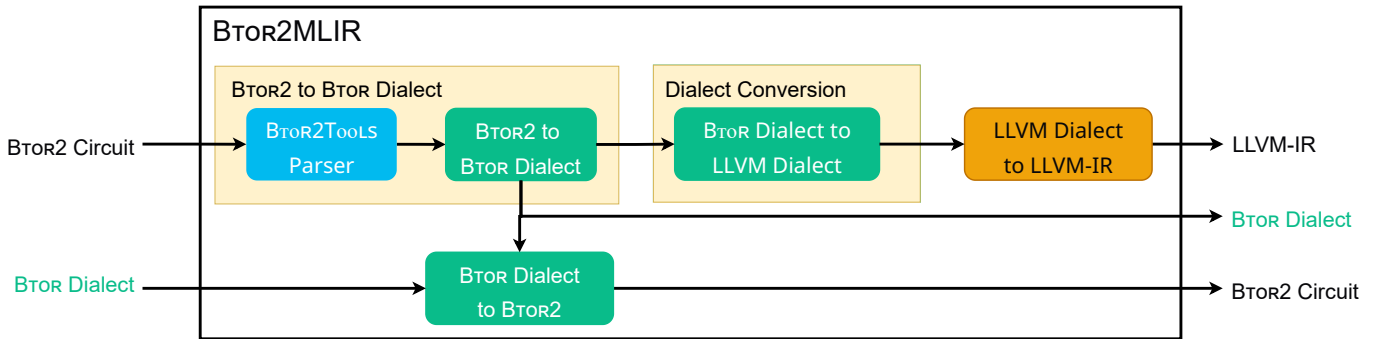


Fig. 2: BTOR2MLIR Architecture.

is `mul`, and `btor.slice` is `slice`. Safety properties such as `bad` are represented by `btor.assert_not`. Special operators such as `one`, `ones` and `constd` are represented by the `btor.constant` operation with the expected integer value. Boolean operators are represented by `btor.cmp`. For example, `eq` becomes `btor.cmp eq`.

Translating BTOR2 to BTOR DIALECT: BTOR2MLIR takes BTOR2 circuits as input, using BTOR2TOOLS to create a data structure for each BTOR2 line. Our pass then generates a program in BTOR DIALECT by constructing the appropriate MLIR AST. Each BTOR2 operator is mapped to a unique operation in BTOR DIALECT, a capability that is greatly simplified and enabled by the MLIR infrastructure.

A program in MLIR can be written using multiple dialects since the MLIR framework enables the interaction of multiple IRs. To enable this capability, MLIR provides dialects that are designed to serve as building blocks for more domain-specific dialects. We utilized the framework by building the BTOR DIALECT using the STANDARD and BUILTIN dialects. For example, we use the `module`, `func` and `bb` operations from BUILTIN. We utilize the `br` operation in the STANDARD dialect to enable interaction between the two basic blocks in Fig. 3b. This approach is consistent with the intended use of the STANDARD and BUILTIN dialects. It saves time and effort since we do not need to recreate operations that already exist in other dialects. Furthermore, MLIR provides a conversion pass from STANDARD dialect to LLVM dialect, making it worthwhile to build BTOR DIALECT on top of the BUILTIN

and STANDARD dialects.

Dialect conversion: The BTOR2MLIR conversion pass from BTOR DIALECT to LLVM dialect utilizes the MLIR infrastructure for pattern-based rewrites. It rewrites BTOR DIALECT operations into LLVM dialect operations. For most operations in BTOR DIALECT there exists a semantically equivalent operation in LLVM dialect. For example, `btor.constant` in Fig. 3b is converted to `llvm.mlir.constant` in LLVM dialect. For some operations, an equivalent in LLVM dialect does not exist, in these cases it is required to rewrite them into several LLVM operations (e.g., in `btor.slice`) and/or to modify the module structure (e.g., `btor.assert_not`). In LLVM dialect, `btor.slice` is replaced by a logical shift right, `llvm.lshr`, and a truncation operation, `llvm.trunc`. `btor.assert_not` is mapped to a new basic block in the LLVM dialect that has the `llvm.unreachable` operation. We split the basic block `bb1`, in Fig. 3b, by adding a conditional branch, `llvm.cond_br`, to direct control flow to the new block when the assertion is satisfied.

Translate LLVM Dialect to LLVM-IR: BTOR2MLIR uses a translation pass from LLVM dialect to LLVM-IR, provided by MLIR. Note the optimizations in the resulting LLVM-IR, shown in Fig. 3c, such as constant propagation and phi nodes.

IV. CORRECTNESS

When introducing a new tool or framework to the community, there is always a question of how polished it is. BTOR2MLIR builds on two mature frameworks:

```

1 sort bitvec 4
2 one 1
3 state 1 factorial
4 state 1 i
5 init 1 3 2
6 init 1 4 2
7 add 1 4 2
8 mul 1 3 4
9 next 1 4 7
10 next 1 3 8
11 ones 1
12 sort bitvec 1
13 eq 12 4 11
14 bad 13
15 slice 12 3 0 0
16 constd 1 3
17 ugt 12 4 16
18 and 12 17 15
19 bad 18

```

(a) Factorial in BTOR2.

```

module {
func @main() {
  %0 = btor.constant 1 : i4
  br ^bb1(%0, %0 : i4, i4)
^bb1(%1: i4, %2: i4):
  %3 = btor.constant 1 : i4
  %4 = btor.add %2, %3 : i4
  %5 = btor.mul %1, %2 : i4
  %6 = btor.constant -1 : i4
  %7 = btor.cmp eq, %2, %6 : i4
  btor.assert_not(%7)
  %8 = btor.constant 0 : i4
  %9 = btor.constant 0 : i4
  %10 = btor.slice %1, %8, %9 : i4,
        i1
  %11 = btor.constant 3 : i4
  %12 = btor.cmp ugt, %2, %11 : i4
  %13 = btor.and %12, %10 : i1
  btor.assert_not(%13)
  br ^bb1(%5, %4 : i4, i4)
}
}

```

(b) Factorial in BTOR DIALECT.

```

declare void @__VERIFIER_error()
define void @main() !dbg !3 {
br label %1
1:
; preds = %14, %0
%2 = phi i4 [%5,%14], [1,%0]
%3 = phi i4 [%4,%14], [1,%0]
%4 = add i4 %3, 1
%5 = mul i4 %2, %3
%6 = icmp eq i4 %3, -1
%7 = xor i1 %6, true
br i1 %7, label %8, label %15
8:
; preds = %1
%9 = lshr i4 %2, 0
%10 = trunc i4 %9 to i1
%11 = icmp ugt i4 %3, 3
%12 = and i1 %11, %10
%13 = xor i1 %12, true
br i1 %13, label %14, label %16
14:
; preds = %8
br label %1
15:
; preds = %1
call void @__VERIFIER_error()
unreachable
16:
; preds = %8
call void @__VERIFIER_error()
unreachable
}

```

(c) Factorial in LLVM-IR.

Fig. 3: BTOR2 to BTOR DIALECT.

	original			roundtrip		
	time	safe/unsafe	TO	time	safe/unsafe	TO
bitvectors						
wolf/18D	157	34/0	2	168	34/0	2
wolf/19A	146	0/1	17	151	0/1	17
wolf/19B	2	3/0	0	2	3/0	0
wolf/19C	834	108/0	5	797	108/0	5
19/beem	278	9/2	4	280	10/2	3
19/goel	190	26/2	43	176	26/2	43
19/mann	4442	29/15	9	4751	30/15	8
20/mann	257	10/5	0	268	10/5	0
bitvectors + arrays						
wolf/18A	70	20/0	0	71	20/0	0
wolf/19B	2	2/3	0	2	2/3	0
19/mann	126	1/1	1	138	1/1	1
20/mann	18	3/3	0	18	3/3	0

TABLE I: Comparing round tripped files.

BTOR2TOOLS and MLIR. This is done not only because of the frameworks’ functionalities, but because they have been extensively reviewed, used, and tested. BTOR2TOOLS has been widely used in the hardware model-checking community since its introduction in 2018. MLIR builds on LLVM, a compiler framework that has been used and improved over numerous projects in the last two decades and is actively supported by industry.

Specifically, BTOR2MLIR uses the parser from BTOR2TOOLS to generate corresponding operations and functions in the BTOR DIALECT of MLIR. The BTOR DIALECT is written in TABLEGEN— an MLIR domain-

specific language for dialect creation. We show how our dialect and the class of binary operations are defined in Fig. 4a. For example, the `BtorBinaryOp` class defines a class of operations that have two arguments `lhs`, `rhs` and a result `res`. It also has a trait `SameOperandsAndResultType` to enforce that `lhs`, `rhs` and `res` have the same type. Finally, the class specifies how the default MLIR parsers and printers should handle such operations. We create our arithmetic operations as shown in Fig. 4b. We mark relevant operations as `Commutative`. Operation descriptions are not shown for simplicity. We ensure that each BTOR2 operator has a one-to-one mapping with an operation in the BTOR DIALECT so that the translation from BTOR2 to BTOR DIALECT is lossless and preserves BTOR2 semantics.

BTOR2MLIR relies on the optimization, folding, and canonicalization passes that MLIR provides in its translation from the LLVM Dialect in MLIR to LLVM-IR. MLIR also provides the mechanism for pattern-based rewrites which has helped us avoid the introduction of undefined behavior into the resulting LLVM-IR. We show an example of this in Fig. 5. MLIR allows us to identify which operations in the BTOR DIALECT we want to replace at the end of our conversion pass. A subset of such operations are shown in Fig. 5a. For each operation that has been identified, we provide a lowering that maps it to a legal operation in the LLVM dialect. We are able to use lowering patterns like `VectorConvertToLLVMPattern` from MLIR for common arithmetic and logical operations as shown in Fig. 5b.

We performed extensive testing using the HWMCC20 benchmark set to verify the correctness of BTOR2MLIR. This is the same benchmark set used to test [15]. The tests are run on a Linux machine with `x86_64` architecture, using BTORMC

```

def Btor_Dialect : Dialect {
  ...
}

class BtorArithmeticOp<string mnemonic, list<Trait> traits = []> :
  Op<Btor_Dialect, mnemonic, traits>;

class BtorBinaryOp<string mnemonic, list<Trait> traits = []> :
  BtorArithmeticOp<mnemonic, !listconcat(traits
    [SameOperandsAndResultType])>,
  Arguments<(ins SignlessIntegerLike:$lhs,
    SignlessIntegerLike:$rhs)>,
  Results<(outs SignlessIntegerLike:$result)>
  {
    let assemblyFormat = "$lhs `,' $rhs attr-dict `:' type($result)";
  }
}

```

(a) Creating BTOR DIALECT.

```

def AddOp : BtorBinaryOp<"add",
  [Commutative]> {
  ...
}

def SubOp : BtorBinaryOp<"sub"> {
  ...
}

def MulOp : BtorBinaryOp<"mul",
  [Commutative]> {
  ...
}

def UDivOp : BtorBinaryOp<"udiv"> {
  ...
}

```

(b) Creating Operations for BTOR DIALECT.

Fig. 4: Using TABLEGEN for Dialect Creation.

with an unroll bound of 20, a timeout of 300 seconds and memory limit of 65 GB. We present the results in Table I, where bitvector benchmarks categories are in the top half and bitvector + array benchmark categories are in the bottom half. All times in this table reflect solved instances and do not include timeouts. We do not show the time it takes to run BTOR2MLIR since the time is negligible. The results are grouped by competition contributor such that each row shows the time, instances solved (safe/unsafe) and timeouts (TO) for both the original and round-tripped circuits. For example, for the wolf/18D category, we can see that the original BTOR2 circuit solves 34 safe instances and 0 unsafe instances in 157 seconds, with 2 timeouts. The round-tripped circuit solves 34 safe instances and 0 unsafe instances in 168 seconds with two timeouts.

We can see that the safety properties in BTOR2 circuits are neither changed nor violated after being round-tripped by BTOR2MLIR. In two categories with only bitvectors, 19/beem and 19/mann, one more instance in each category is found safe after round trip, while the original circuit leads to a memout and timeout respectively. This gives us confidence that the translation to BTOR DIALECT, using the BTOR2TOOLS parser, is indeed correct. Then, we tested whether the same holds after translation to LLVM-IR. Through this method, we were able to ensure that BTOR2MLIR does not have errors when handling operations that are represented in the benchmark set. This approach is not complete, however, since it would not identify errors that might be in our implementation but are not exercised by the benchmarks we use. For example, BTOR2 expects that a division by zero would result in -1 , but there are no benchmarks that exercise this kind of division. We mitigate this by generating benchmarks for division, remainder, and modulus operators to ensure that the expected behavior of BTOR2 operators are represented in our test suite.

In the future, it is interesting to explore other translation validation and verification approaches. For example, it would be useful for BTOR2MLIR to produce a proof trail that

justifies all of the transformations that are performed by the tool. This, for example, might be possible to achieve by building on the work of [16], [17].

Limitations: BTOR2MLIR is able to round trip BTOR2 operators and their sorts. In LLVM-IR all BTOR2 operators and their sorts are supported as well, but not fairness and justice constraints.

V. EVALUATION

To evaluate BTOR2MLIR, we have built a prototype hardware model checker by connecting our tool with SEAHORN [8], a well-known model checker for C/C++ programs that works at the LLVM-IR level. It has recently been extended with a bit-precise Bounded Model Checking engine [18]. This BMC engine was evaluated in a recent case study [19] and we use the same configuration of SEAHORN in our evaluation.

The goal of our evaluation is to show that BTOR2MLIR makes it easy to connect hardware designs with LLVM-based verification engines. We did not expect the existing software engines to outperform dedicated hardware model checkers. However, we hope that this will enable further avenues of research. In the future, we plan to extend the framework to support other LLVM-based analysis tools, such as symbolic execution engine KLEE [7], and fuzzing framework [6].

For the evaluation, we have chosen the bitvector category of BTOR benchmarks from the most recent Hardware Model Checking Competition (HWMCC) [2]. We have excluded benchmarks with arrays since the export to LLVM-IR is not supported by SEAHORN in our experimental setup. All our experiments are run on a Linux machine with x86_64 architecture, with unroll bound of 20, a timeout of 300 seconds and memory limit of 65 GB. The results are presented in Table II, grouped by competition contributor. All times in this table reflect solved instances and do not include timeouts. We do not show the time it takes to run BTOR2MLIR since the time is negligible. In the rest of this section, we highlight some of the interesting findings.

We have run BTORMC on the same machine and exact same experimental setup (unroll bound and CPU and memory


```

void BtorToLLVMLoweringPass::runOnOperation() {
  LLVMConversionTarget target(getContext());
  RewritePatternSet patterns(getContext());
  LLVMTypeConverter converter(&getContext());
  mlir::btor::populateBtorToLLVMConversionPatterns(converter,
    patterns);
  ...
  /// binary operators
  // arithmetic
  target.addIllegalOp<btor::AddOp, btor::SubOp, btor::MulOp,
    btor::UDivOp...>();
  ...
}

```

(a) Identifying operations.

```

...
using AddOpLowering =
  VectorConvertToLLVMPattern<btor::AddOp, LLVM::AddOp>;
using SubOpLowering =
  VectorConvertToLLVMPattern<btor::SubOp, LLVM::SubOp>;
using MulOpLowering =
  VectorConvertToLLVMPattern<btor::MulOp, LLVM::MulOp>;
using UDivOpLowering =
  VectorConvertToLLVMPattern<btor::UDivOp, LLVM::UDivOp>;
...
void mlir::btor::populateBtorToLLVMConversionPatterns(
  LLVMTypeConverter &converter, RewritePatternSet
  &patterns) {
  patterns.add<
    AddOpLowering, SubOpLowering, MulOpLowering,
    UDivOpLowering, ...>(converter);
}
...

```

(b) Converting operations to LLVM-IR.

Fig. 5: Using Patter Based Rewriters in MLIR.

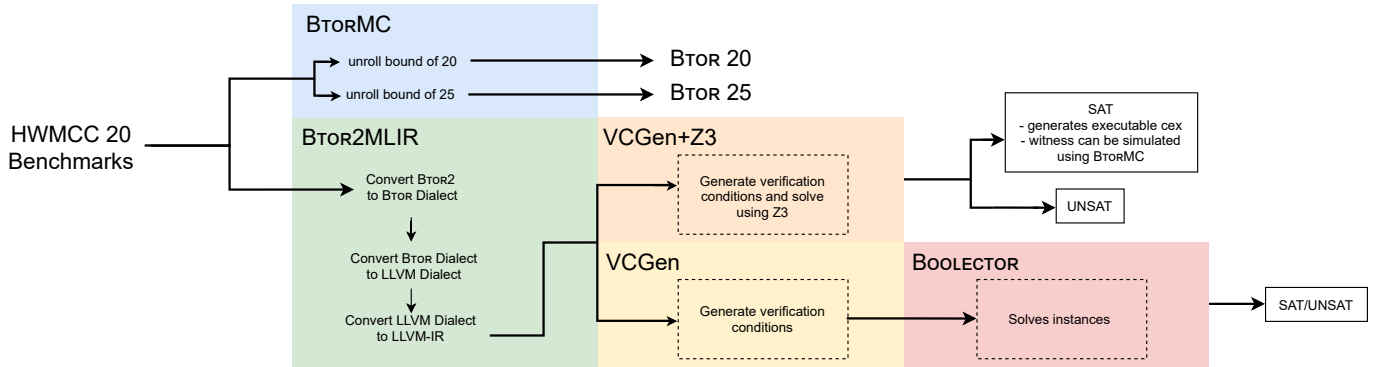


Fig. 6: Verification Strategies.

limits). We chose BTORMC because it is well integrated with the HWMCC environment and is specifically designed for BTOR2. The results of running BTORMC are shown in the first columns of BTORMC in Table II. For each category, we show the total time for all instances that are solved in that category, and the number of instances that are solved as safe, unsafe, and timed-out (TO), respectively. For example, the 20/mann category is solved in 257 seconds, 10 instances are safe, 5 are unsafe, and no instance has timed out. The performance of BTORMC is quite good across the board.

We evaluate the problems generated by BTOR2MLIR by

plugging them into SEAHORN. SEAHORN pre-processes programs before attempting to verify them. This includes, standard LLVM optimizations (i.e., -O3), loop unrolling and loop cutting are applied. We found that SEAHORN was able to, in some instances, remove the assertions in the LLVM-IR, meaning that the program was found to be safe statically, before invoking the BMC. The BMC also runs simplifications on the formulas that it sends to Z3, its default underlying SMT solver. The results for this run are shown in the Z3 columns of Table II. For example, the 20/mann category is solved in 94 seconds, 8 instances are safe, 5 are unsafe and 2 have

		BTORMC		SEAHORN					BTORMC		SEAHORN		
		20	25	VCGen + Z3	VCGen	BTOR			20	25	VCGen + Z3	VCGen	BTOR
wolf/18D	Time (s)	157	394	560	543	745	19/beam	Time (s)	278	251	309	35	85
	Safe	34	34	29	-	34		Safe	9	8	6	-	7
	Unsafe	0	0	0	-	0		Unsafe	2	2	2	-	2
	TO	2	2	7	2	2		TO	4	5	7	4	6
wolf/19A	Time (s)	146	106	-	-	-	19/goel	Time (s)	190	349	489	132	335
	Safe	0	0	0	-	0		Safe	26	25	25	-	28
	Unsafe	1	1	0	-	0		Unsafe	2	2	1	-	2
	TO	17	17	18	18	18		TO	43	44	45	27	41
wolf/19B	Time (s)	2	2	2	2	3	19/mann	Time (s)	4442	8674	3811	175	3015
	Safe	3	3	3	-	3		Safe	29	28	19	-	30
	Unsafe	0	0	0	-	0		Unsafe	15	15	14	-	14
	TO	0	0	0	0	0		TO	9	10	20	2	9
wolf/19C	Time (s)	834	1101	354	418	1085	20/mann	Time (s)	257	495	94	35	188
	Safe	108	107	102	-	106		Safe	10	10	8	-	9
	Unsafe	0	0	0	-	0		Unsafe	5	5	5	-	5
	TO	5	6	11	2	7		TO	0	0	2	0	1

TABLE II: HWMCC20 Results.

timed out. The reported time does not include the instances that have timed out.

The aggregate time of SEAHORN on most of the categories is higher than that of BTORMC, often by a significant amount. We looked into this and found that SEAHORN treats the given bound as a lower bound, rather than an upper bound. That is, it ensures that it unrolls the programs to a depth of at least 20, but it may continue past that point. Taking this into account, we ran BTORMC with a bound of 25. The results are in the second columns of BTORMC in Table II. As expected, its aggregate times are higher than the run of BTORMC with bound 20. We notice, however, that it is slower than SEAHORN in the 19/mann category.

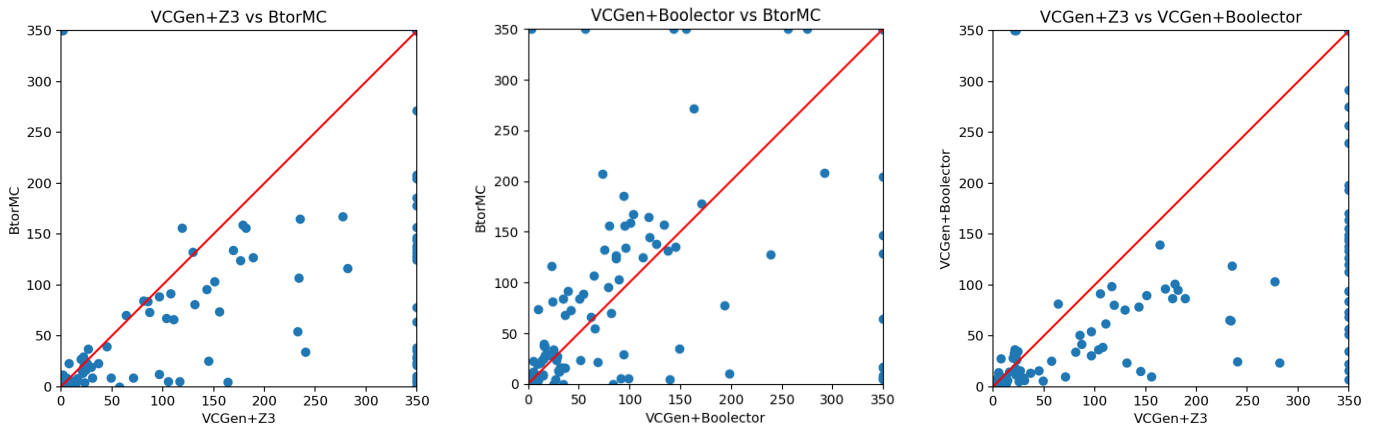
BOOLECTOR and Z3 are the SMT solvers used by BTORMC and SEAHORN respectively. Given that BOOLECTOR is optimized for BTOR2 circuits, we evaluated whether the SMT formulas generated by SEAHORN would be solved faster by BOOLECTOR. The results for generating SMT-LIB formulas using SEAHORN are presented in the VCGen column of Table II. The times are low for most categories except wolf/18D, wolf/19C, 19/goel and 19/mann. For example, it takes SEAHORN 175 seconds to generate the verification conditions for instances in the 19/mann category, with two timeouts. This includes the time it takes SEAHORN to print the SMT formulas to disk. We plug the resulting SMT formulas into BOOLECTOR and present the results in the BTOR columns of Table II. The results show that using SEAHORN to generate verification conditions and BOOLECTOR to solve these instances is often better than using BTORMC. For example, for category 19/mann, it takes 3015s for BOOLECTOR to solve 44 instances with 9 timeouts. Therefore, the total time for SEAHORN and BOOLECTOR (3190) represents the

	BTORMC		SEAHORN		
	20	25	VCGen+Z3	VCGen	BTOR
Time (s)	6309	11373	5621	1340	5456
Safe	219	215	192	-	217
Unsafe	25	25	22	-	23
TO	80	84	110	55	84

TABLE III: Total results for each tool.

time it takes to translate, generate SMT formula and verify the 19/mann category. Note that two of the 9 timeouts in this category are attributed to the fact that SEAHORN has a timeout when generating verification conditions.

To get the big picture of how the different infrastructures performed, we collected the results over all categories in Table III. From this table, we can see that our hybrid pipeline combining BTOR2MLIR, SEAHORN, and BOOLECTOR solves 240 instances with 84 timeouts in 6796s (sum of VCGen and BTOR total times), which is very encouraging. We also present plots that compare the different pipelines that have been explored in Fig. 7. We set the time for all timeout instances to 350 seconds so that they are distinguished from instances that were solved close to the timeout threshold. First, we look at the performance of the hybrid pipeline that combines BTOR2MLIR, SEAHORN and its default SMT solver Z3 against BTORMC in Fig. 7a. Z3 does as well as BTORMC for most instances that are easy, however, it struggles when the problems are harder. This is not as clear from Table II since focuses on the number of timeouts and benchmarks solved. Second, we present the performance of BTORMC against the hybrid pipeline that combines BTOR2MLIR, SEAHORN and



(a) VCGen + Z3 vs BTORMC. (b) VCGen + BOOLECTOR vs BTORMC. (c) VCGen + Z3 vs VCGen + BOOLECTOR.

Fig. 7: Verification strategy comparison.

BOOLECTOR in Fig. 7b. We can see that there are more benchmarks that this pipeline solves faster than BTORMC. It is also clear that it solves more benchmarks than the Z3 configuration in Fig. 7a, as we would expect from Table III. Third, we compare the two hybrid pipelines in Fig. 7c. We can see that the configuration that uses SEAHORN to generate verification conditions and BOOLECTOR for solving easily outperforms the Z3 configuration.

VI. RELATED WORK

Translating BTOR2 circuits into other formats enables the application of different verification methods and techniques. The gains that can be made from applying one method of encoding over another could enable solving a class of benchmarks that are not solved with existing approaches.

BTOR2LLVM⁶ and BTOR2CHC⁷ are tools that convert BTOR2 circuits to programs in LLVM-IR and CHCs, respectively. These tools are developed in Python, in order to be light weight, but end up repeating shared functionality and tools since they lack a common infrastructure. Translated BTOR2 benchmarks⁸ have also been collected to facilitate research, but information of what tools were used to get the CHC format is not publicly available. While a collection of translated benchmarks is valuable, it is important that there are tools to do the translation on demand. This enables rapid prototyping in a way that saved benchmarks do not.

BTOR2C [15] is a recent tool that converts BTOR2 circuits to C programs. It has been used to facilitate the utilization of software analyzers by serving as a pre-processing step that bridges the gap between the world of software verification and hardware verification. There are limitations that arise, however, from differences in the semantics of BTOR2 and

C. An important limitation that C imposes on this project is the inability to represent arbitrary width bitvectors. This means that BTOR2 circuits which operate on bitvectors of width greater than 128 are not supported. These limitations, as well as BTOR2C lack of support for BTOR2 operators that have overflow detection are resolved by using LLVM-IR as the target language.

A common theme across these efforts is that they are not built on an architecture that can be easily extended. Each project aims to make it easier to utilize advances in formal verification, but they fail to offer a solution that does not require recreating components that already exist.

VII. CONCLUSION

In this paper, we present BTOR2MLIR — a new format and toolchain for hardware verification, based on the MLIR intermediate representation framework of the LLVM compiler infrastructure. Our goal is to open new doors for the research and applications of hardware verification by taking advantage of recent innovations in compiler construction technology. We believe that this project opens new avenues for exploring the application of existing verification and testing techniques developed for software to the hardware domain. As a proof of concept, we have connected BTOR2MLIR with the SEAHORN verification engine. While out-of-the-box, this gives acceptable performance, when combined with BOOLECTOR, a combination that is competitive against BTORMC. In the future, we plan to continue this line of research and explore applying testing and simulation technologies such as KLEE [7] and LIBFUZZER [6]. We also plan to generate formats for other verification techniques such as AIGER [20], Constrained Horn Clauses, and SMT-LIB.

REFERENCES

- [1] S. Malik, “Hardware verification: Techniques, methodology and solutions,” in *Tools and Algorithms for the Construction and Analysis of*

⁶<https://github.com/stepwise-alan/btor2llvm>

⁷<https://github.com/stepwise-alan/btor2chc>

⁸<https://github.com/zhanghongce/HWMCC19-in-CHC>

- Systems, C. R. Ramakrishnan and J. Rehof, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 1–1.
- [2] A. Biere, T. van Dijk, and K. Heljanko, “Hardware model checking competition 2017,” in *2017 Formal Methods in Computer Aided Design (FMCAD)*, 2017, pp. 9–9.
- [3] A. Niemetz, M. Preiner, C. Wolf, and A. Biere, “Btor2 , BtorMC and Boolector 3.0,” in *Computer Aided Verification*, H. Chockler and G. Weissenbacher, Eds. Cham: Springer International Publishing, 2018, pp. 587–595.
- [4] C. Lattner and V. Adve, “LLVM: a compilation framework for lifelong program analysis & transformation,” in *International Symposium on Code Generation and Optimization, 2004. CGO 2004.*, 2004, pp. 75–86.
- [5] C. Lattner, M. Amini, U. Bondhugula, A. Cohen, A. Davis, J. Pienaar, R. Riddle, T. Shpeisman, N. Vasilache, and O. Zinenko, “MLIR: A Compiler Infrastructure for the End of Moore’s Law,” 2020.
- [6] K. Serebryany, “Continuous Fuzzing with libFuzzer and AddressSanitizer,” in *2016 IEEE Cybersecurity Development (SecDev)*, 2016, pp. 157–157.
- [7] C. Cadar, D. Dunbar, and D. R. Engler, “Klee: Unassisted and automatic generation of high-coverage tests for complex systems programs,” in *USENIX Symposium on Operating Systems Design and Implementation*, 2008.
- [8] A. Gurfinkel, T. Kahsai, A. Komuravelli, and J. A. Navas, “The SeaHorn Verification Framework,” in *Computer Aided Verification*, D. Kroening and C. S. Păsăreanu, Eds. Cham: Springer International Publishing, 2015, pp. 343–361.
- [9] A. Gurfinkel and J. A. Navas, “Abstract interpretation of LLVM with a region-based memory model,” in *Software Verification - 13th International Conference, VSTTE 2021, New Haven, CT, USA, October 18-19, 2021, and 14th International Workshop, NSV 2021, Los Angeles, CA, USA, July 18-19, 2021, Revised Selected Papers*, ser. Lecture Notes in Computer Science, R. Bloem, R. Dimitrova, C. Fan, and N. Sharygina, Eds., vol. 13124. Springer, 2021, pp. 122–144. [Online]. Available: https://doi.org/10.1007/978-3-030-95561-8_8
- [10] S. Eldridge, P. Barua, A. Chapyzenka, A. Izraelevitz, J. Koenig, C. Lattner, A. Lenharth, G. Leontiev, F. Schuiki, R. Sunder, A. Young, and R. Xia, “MLIR as Hardware Compiler Infrastructure,” in *Workshop on Open-Source EDA Technology (WOSET)*, 2021.
- [11] C. Barrett, P. Fontaine, and C. Tinelli, “The Satisfiability Modulo Theories Library (SMT-LIB),” www.SMT-LIB.org, 2016.
- [12] C. Wolf, “Yosys open synthesis suite,” <https://yosyshq.net/yosys/>.
- [13] S. Palnitkar, *Verilog HDL: A Guide to Digital Design and Synthesis*. USA: Prentice-Hall, Inc., 1996.
- [14] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. A. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: A system for large-scale machine learning,” in *12th USENIX Symposium on Operating Systems Design and Implementation, OSDI 2016, Savannah, GA, USA, November 2-4, 2016*, K. Keeton and T. Roscoe, Eds. USENIX Association, 2016, pp. 265–283. [Online]. Available: <https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi>
- [15] D. Beyer, P.-C. Chien, and N.-Z. Lee, “Bridging hardware and software analysis with Btor2C: A word-level-circuit-to-C translator,” in *Proc. TACAS*, ser. LNCS 13994. Springer, 2023, pp. 1–21. [Online]. Available: <https://www.sosy-lab.org/research/btor2c/>
- [16] S. Chatterjee, A. Mishchenko, R. K. Brayton, and A. Kuehlmann, “On resolution proofs for combinational equivalence,” in *Proceedings of the 44th Design Automation Conference, DAC 2007, San Diego, CA, USA, June 4-8, 2007*. IEEE, 2007, pp. 600–605. [Online]. Available: <https://doi.org/10.1145/1278480.1278631>
- [17] R. E. Bryant, “Tbuddy: A proof-generating BDD package,” in *22nd Formal Methods in Computer-Aided Design, FMCAD 2022, Trento, Italy, October 17-21, 2022*, A. Griggio and N. Rungta, Eds. IEEE, 2022, pp. 49–58. [Online]. Available: https://doi.org/10.34727/2022/isbn.978-3-85448-053-2_10
- [18] S. Priya, X. Zhou, Y. Su, Y. Vizel, Y. Bao, and A. Gurfinkel, “Bounded Model Checking for LLVM,” in *Formal Methods in Computer Aided Design, FMCAD 2022*, 2022, p. 214.
- [19] —, “Verifying verified code,” *Innov. Syst. Softw. Eng.*, vol. 18, no. 3, pp. 335–346, 2022. [Online]. Available: <https://doi.org/10.1007/s11334-022-00443-9>
- [20] A. Biere, K. Heljanko, and S. Wieringa, “AIGER 1.9 and beyond,” Institute for Formal Models and Verification, Johannes Kepler University, Altenbergerstr. 69, 4040 Linz, Austria, Tech. Rep. 11/2, 2011.