

Array representations for Polly

(Anonymous)

August 21, 2018

Contents

1	Introduction	3
2	Formalization	3
3	C-style arrays — no stride, no offset	4
4	Fortran-style arrays: only offset	4
5	Chapel-style arrays: strides and offsets	4
6	Caveats	5
7	References	5

1 Introduction

We mathematically describe the array representations that are used by Fortran and Chapel, two programming languages which support complex array indexing. We motivate the need for LLVM to be able to represent these array representations and their coresponding indexing operation at a semantic level that is higher than that of `getelementptr`. We showcase existing implementations of such a semantic representation, and would like to propose this as an LLVM standard. Choices for implementation are outlined in this document.

2 Formalization

We first formalize the most general type of array we have encountered - That of Chapel style arrays which are multidimensional, strided, and provide non-zero-based indexing.

We denote the set of all possible arrays (a loosely defined term, to be sure) as *Array*.

$$dim : Array \rightarrow \mathbb{N}$$

$$dim(A) = \text{dimensionality of array } A$$

$$ixbegin/ixend : Array \times [0, dim(A)] \rightarrow \mathbb{Z}$$

$$ixbegin/ixend(A, d) = \text{first and last legal indeces allowed along dimension } d$$

$$stride : Array \times [0, dim(A)] \rightarrow \mathbb{N}$$

$$stride(A, d) = \text{stride of an array } A \text{ along dimension } d$$

Using the data outlined above, we can now define what a correct indexing function into an array is. An indexing function recieves **logical coordinates** of the array index, which is converted into **physical coordinates**, which is the in-memory representation.

We now try to define a partial function *logicalToPhysical*, which maps logical coordinates to physical coordinates.

First, we define *ixset*(*A*, *d*), the set of valid indeces along a dimension *d*:

$$ixset : Array \times [0, dim(A)] \rightarrow 2^{\mathbb{N}}$$

$$ixset(A, d) = \{ixbegin(A, d) + \alpha \cdot stride(A, d) \mid \alpha \in \mathbb{N}\} \cap (-\infty, ixend(A, d)]$$

The *ixset* constructs the set of valid indeces that are allowed along a given dimension. We first give the type of *logicalToPhysical*, and we then gradually build up the full expression.

$$\text{logicalToPhysical} : \text{Array} \times \prod_{d=1}^{\dim(A)} \text{ixset}(A, d) \rightarrow \mathbb{N}$$

Notice that the logical coordinates live in $\prod_{d=1}^{\dim(A)} \text{ixset}(A, d)$, while physical coordinates live in \mathbb{N} . We wish to view memory as a 1-D sequence of values, with 0 being the "base address" of the array. The logical coordinates allow us to index the "virtual view" of the array that we possess.

3 C-style arrays — no stride, no offset

In the case of C-style arrays, our $\text{ixset}(A, i) = [0, N(A, i)) \subset \mathbb{N}$, where $N(A, i)$ is informally the "size" of array A along dimension i . The word "size" is, however, fraught with danger, since it conflates the size of the "logical indexing space" and the underlying "memory space". In the case of the C language, both of these concepts coincide, but this is not so in other languages such as Chapel.

In the case of a C array, here are the different functions that we have defined:

- Declaration: $\text{T } A[N_1][N_2] \dots [N_d]$
- Dimension: $\dim(A) = d$
- Beginning, ending indices: $\text{ixbegin}(A, _) = 0$, $\text{ixend}(A, d) = N_d - 1$
- Stride: $\text{stride}(A, _) = 1$
- Index set:
$$\begin{aligned} \text{ixset}(A, d) &= \{\text{ixbegin}(A, d) + \alpha \cdot \text{stride}(A, d) \mid \alpha \in \mathbb{N}\} \cap (-\infty, \text{ixend}(A, d)] \\ &= \{0 + \alpha \cdot 1 \mid \alpha \in \mathbb{N}\} \cap (-\infty, N_d - 1] \\ &= [0, \infty) \cap (-\infty, N_d - 1] \\ &= [0, N_d - 1] \end{aligned}$$

To calculate the index expression, we create a function, $\text{volume}(A, d) = |xx|$

4 Fortran-style arrays: only offset

5 Chapel-style arrays: strides and offsets

$$\text{logicalToPhysical}(A, (ix_1, ix_2, \dots, ix_{\dim(A)})) = \sum_{i=1}^{\dim(A)} (ix_i / \text{stride}(A, i) + \text{offset}(A, i))$$

6 Caveats

We do not take into account alignment, which is something that must be done for this document to be complete.

7 References

- The Chapel language specification.
- Fortran 2003 standard.