Array Types for a Graph Processing Language

Mark Tullsen and Matt Sottile

Galois, Inc.

GABB, May 2016

Outline

- Introduction
- Capturing Structure
- Representations
- (I.e., Maps, Dictionaries, Etc.)
- Index Spaces
- Larger Examples
- Conclusion



- Introduction
- Capturing Structure
- 3 Representations
- 4 "Partial Arrays" (I.e., Maps, Dictionaries, Etc.)
- Index Spaces
- 6 Larger Examples
- Conclusion

LPG and HAL

The context of HAL: designing & implementing LPG

LPG (Language for Processing Graphs):

- A declarative DSL for medium-size¹ graph processing.
- Goal: Architecture Neutral
- Goal: Maximize implicit parallelism

HAL (Hierarchical Array Language):

- a declarative array language,
- the primary abstraction between our graph algorithms and the parallel architecture.



¹Fit on single-host.

Why an Intermediate Array Language?

Why an Intermediate Array Language?

- Graph algorithms
 - Need many data structures, not just graphs
 - Often look like array-processing
 - Existing work on parallelizing array languages
- Very natural when thinking in terms of adjacency matrices
 - Classes of Graphs have analogs in HAL's various array structures
- We expect this will enable more efficient code
 - More parallelism
 - Have a richer set of laws (for transformation)

HAL (Hierarchical Array Language)

- Declarative
- A very expressive type system
 - captures more structure
 - provides more laws
 - sums!
 - allows us to provide bijections and views
- Standard and "partial" (associative) arrays
- First class index spaces

HAL (Hierarchical Array Language)

- Declarative
- A very expressive type system
 - captures more structure
 - provides more laws
 - sums!
 - allows us to provide bijections and views
- Standard and "partial" (associative) arrays
- First class index spaces

CAVEAT:

HAL (Hierarchical Array Language)

- Declarative
- A very expressive type system
 - captures more structure
 - provides more laws
 - sums!
 - allows us to provide bijections and views
- Standard and "partial" (associative) arrays
- First class index spaces

CAVEAT:

 HAL is in the design & prototype stage: no compiler and no performance figures yet.

- 1 Introduction
- 2 Capturing Structure
- Representations
- 4 "Partial Arrays" (I.e., Maps, Dictionaries, Etc.)
- Index Spaces
- 6 Larger Examples
- Conclusion

A simple matrix, m

```
m :: \mathbb{Z}_4 \times \mathbb{Z}_4 \Rightarrow \mathbb{Z}
0 0 0 0
                                                       m :: VEC^2(\mathbb{Z}_4) \Rightarrow \mathbb{Z}
0 1 0 0
0 0 2 0
0 0 0 3
m_1 = arr[\mathbb{Z}_4 \times \mathbb{Z}_4] [0 0 0 0
                                                                                |m_1| = 1+4\times 4
                               0 0 0 31
m_2 = unnest(arr[\mathbb{Z}_4])
                                                                                |m_2| = 1+4\times4
                              [arr[\mathbb{Z}_4] [0 0 0 0]
                              , arr[\mathbb{Z}_4] [0 1 0 0]
                              , arr[\mathbb{Z}_4] [0 0 2 0]
                              , arr[\mathbb{Z}_4] [0 0 0 3]
                              1)
```

|m| = storage requirements of m, in machine words

Upper triangular, all zeros

$$u_1 = arr[SET^2(\mathbb{Z}_4)]$$
 [0 0 0 0 0 0] $|u_1| = 1+6$ $u_2 = const[SET^2(\mathbb{Z}_4)]$ 0 $|u_2| = 1$

Upper triangular, all zeros

0 0 0
$$u :: SET^2(\mathbb{Z}_4) \Rightarrow \mathbb{Z}$$
 0 0 0

$VEC^2(\mathbb{Z}_A)$

(0,0)	(0,1)	(0,2)	(0,3)
(1,0)	(1,1)	(1,2)	(1,3)
(2,0)	(2,1)	(2,2)	(2,3)
(3,0)	(3,1)	(3,2)	(3,3)

Upper triangular, all zeros

0 0 0
$$u :: SET^2(\mathbb{Z}_4) \Rightarrow \mathbb{Z}$$
 0 0 0

$$u_1 = arr[SET^2(\mathbb{Z}_4)]$$
 [0 0 0 0 0 0] $|u_1| = 1+6$ $u_2 = const[SET^2(\mathbb{Z}_4)]$ 0 $|u_2| = 1$

$$SET^2(\mathbb{Z}_4)$$

{0,1}	{0,2}	{0,3}
	{1,2}	{1,3}
		{2,3}

A sequence on the diagonal

Combining . . . to define *m* in HAL

```
m :: VEC^2(\mathbb{Z}_4) \Rightarrow \mathbb{Z}
0 0 2 0
0 0 0 3
m_1 = arr[\mathbb{Z}_4 \times \mathbb{Z}_4] [0 \ 0 \ 0 \ 0 \ 1 \dots]
                                                                      |m_1| = 1+4\times4
m_3 = \text{fromTris} (\text{const}[SET^2(\mathbb{Z}_4)] 0
                                                                      | m_3 | = 1 + 1 + 1 + 1
                          , smart [\mathbb{Z}_4] id
                          , const[SET^2(\mathbb{Z}_4)] 0 )
```

- same type, same interface
- const & smart provide
 - correctness by construction
 - smaller representations



- Representations

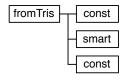
smart/const/... are represented by "tagged values"

```
from Tris (const.., smart.., const..) | | = 1+1+1+1
```

```
fromTris
              const
              smart
              const
```

smart/const/... are represented by "tagged values"

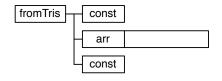
```
from Tris (const.., smart.., const..) | | = 1+1+1+1
```



Could nest arbitrarily, thus "<u>Hierarchical</u> Array Language"

Converting to Expanded Representations

```
from Tris (const.., @smart.., const..) | | = 1+1+4+1
```



- The @ operator expands our const/smart constructors to arr-like, flat representations. Thus, we can define things semantically then get either
 - compact representation
 - flattened representation

Converting to Expanded Representations (2)

```
arr
```

Compact and Expanded Arrays, Mixed

```
m2 :: VEC^2(\mathbb{Z}_4) \Rightarrow \mathbb{Z}
   0 0 0
m2 = fromTris (arr[SET^2(\mathbb{Z}_4)] [5 9 1 2 5 3]
                        , const[\mathbb{Z}_4]
                        , const[SET^2(\mathbb{Z}_4)] 0
```

Compact and Expanded Arrays, Mixed

 NOTE: This is not a substitute for explicitly defining symmetric/triangular matrices.

A Triangular Matrix

5 9 1
$$m3 :: SET^2(\mathbb{Z}_4) \Rightarrow \mathbb{Z}$$
 2 5
$$3$$

$$m3 = arr[SET^2(\mathbb{Z}_4)]$$
 [5 9 1 2 5 3]
 $|m3| = 1 + (4 \text{ choose } 2) = 1 + 6$

- Introduction
- Capturing Structure
- Representations
- 4 "Partial Arrays" (I.e., Maps, Dictionaries, Etc.)
- Index Spaces
- 6 Larger Examples
- Conclusion

Partial Arrays

- Partial Array = map, dictionary, associative array, . . .
- Examples:

```
{1: 2, 5: 3, ...}
                                                      a1 :: \mathbb{Z}_{50} \Longrightarrow \mathbb{Z}
                                                     |a1| = 1+2 \times nnz(a1)
{1: 2, 5: 3, ...} DFLT 0
                                                      a2 :: \mathbb{Z}_{50} \Rightarrow \mathbb{Z}
                                                     |a2| = 1+2 \times nnz(a1)
@({1: 2, 5: 3, ...} DFLT 0)
                                                     a3 :: \mathbb{Z}_{50} \Rightarrow \mathbb{Z}
                                                     |a3| = 1+50
```

nnz - number of non zeros, loosely

Aside: Regarding Sums

A sum type

$$a + b + c$$

- a or b or c (but tagged to distinguish)
- Like 'tagged' unions in C, but safe
- Haskell's "Maybe a" type (ML's "option a") in HAL:

$$\mathbb{Z}_1$$
 + a

Partial Arrays: Semantics

 Semantically we can understand partial arrays through this (provided) bijection between partial and total arrays:

$$a \Longrightarrow b \iff a \Rightarrow \mathbb{Z}_1 + b$$

- Operationally
 - Uses a sparse representation, by default
 - Can iterate over in time proportional to nnz(parray)
- Semantic partiality is orthogonal to sparse representations
 - can have sparse representations for (total) arrays (<u>DFLT</u>)
 - can have non-sparse representations for partial arrays

- **Index Spaces**

Index Space Transformation + Bijection

(0,0)	(0,1)	(0,2)	(0,3)
(1,0)	(1,1)	(1,2)	(1,3)
(2,0)	(2,1)	(2,2)	(2,3)
(3,0)	(3,1)	(3,2)	(3,3)

(0,0)	(0,1)	(0,0)	(0,1)
(1,0)	(1,1)	(1,0)	(1,1)
(0,0)	(0,1)	(0,0)	(0,1)
(1,0)	(1,1)	(1,0)	(1,1)

(0,0)	(0,1)	(0,0)	(0,1)
(1,0)	(1,1)	(1,0)	(1,1)
(0,0)	(0,1)	(0,0)	(0.1)
(0,0)	(0,1)	(0,0)	(0,1)
(1,0)	(1,1)	(1,0)	(1,1)

(0,0)	(0,1)	(0,0)	(0,1)
(1,0)	(1,1)	(1,0)	(1,1)
(0,0)	(0,1)	(0,0)	(0,1)
(1,0)	(1,1)	(1,0)	(1,1)

0	0	0	0	
1	1	1	1	
2	2	2	2	
3	3	3	3	

0	1	2	3
0	1	2	3
0	1	2	3
0	1	2	3

(0,0)	(0,0)	(0,1)	(0,1)
(1,0)	(1,0)	(1,1)	(1,1)
(2,0)	(2,0)	(2,1)	(2,1)
(3,0)	(3,0)	(3,1)	(3,1)

(0,0)	(0,1)	(0,0)	(0,1)
(1,0)	(1,1)	(1,0)	(1,1)
(0,0)	(0,1)	(0,0)	(0,1)
(1,0)	(1,1)	(1,0)	(1,1)

0	0	0
1	1	1
2	2	2
3	3	3
	1 2	1 1 2 2

0	1	2	3
0	1	2	3
0	1	2	3
0	1	2	3

(0,0)	(0,0)	(0,1)	(0,1)
(1,0)	(1,0)	(1,1)	(1,1)
(2,0)	(2,0)	(2,1)	(2,1)
(3,0)	(3,0)	(3,1)	(3,1)

- I.e., if we can partition the index-space (via a bijection)
 - we can decompose the array
- This allows for
 - processor partitioning
 - divide-and-conquer algorithms
 - nice interactions with the smart/const constructors

(0,0)	(0,1)	(0,0)	(0,1)
(1,0)	(1,1)	(1,0)	(1,1)
(0,0)	(0,1)	(0,0)	(0,1)
(1,0)	(1,1)	(1,0)	(1,1)

0	0	0	0	
1	1	1	1	
2	2	2	2	
3	3	3	3	

0	1	2	3
0	1	2	3
0	1	2	3
0	1	2	3

(0,0)	(0,0)	(0,1)	(0,1)
(1,0)	(1,0)	(1,1)	(1,1)
(2,0)	(2,0)	(2,1)	(2,1)
(3,0)	(3,0)	(3,1)	(3,1)

0	{0,1}	{0,2}	{0,3}
{0,1}	1	{1,2}	{1,3}
{0,2}	{1,2}	2	{2,3}
{0,3}	{1,3}	{2,3}	3

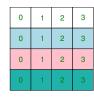
{0,3}	{0,2}	{0,1}	0
{1,3}	{1,2}	1	{0,1}
{2,3}	2	{1,2}	{0,2}
3	{2,3}	{1,3}	{0,3}

{0,0}	{0,1}	{0,2}	{0,3}
{0,1}	{1,1}	{1,2}	{1,3}
{0,2}	{1,2}	{2,2}	{2,3}
{0,3}	{1,3}	{2,3}	{3,3}

0	(0,1)	(0,2)	(0,3)
(1,0)	1	(1,2)	(1,3)
(2,0)	(2,1)	2	(2,3)
(3,0)	(3,1)	(3,2)	3







(0,0)	(0,0)	(0,1)	(0,1)
(1,0)	(1,0)	(1,1)	(1,1)
(2,0)	(2,0)	(2,1)	(2,1)
(3,0)	(3,0)	(3,1)	(3,1)

0	{0,1}	{0,2}	{0,3}
{0,1}	1	{1,2}	{1,3}
{0,2}	{1,2}	2	{2,3}
{0,3}	{1,3}	{2,3}	3

{0,3}	{0,2}	{0,1}	0
{1,3}	{1,2}	1	{0,1}
{2,3}	2	{1,2}	{0,2}
3	{2,3}	{1,3}	{0,3}

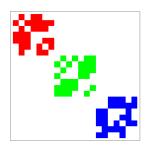
{0,0}	{0,1}	{0,2}	{0,3}	
{0,1}	{1,1}	{1,2}	{1,3}	
{0,2}	{1,2}	{2,2}	{2,3}	
{0,3}	{1,3}	{2,3}	{3,3}	

0	(0,1)	(0,2)	(0,3)
(1,0)	1	(1,2)	(1,3)
(2,0)	(2,1)	2	(2,3)
(3,0)	(3,1)	(3,2)	3

- Without sums, we'd only be able to partition into equal sized parts.
 - The indices remain unique and form a valid type

- Introduction
- Capturing Structure
- Representations
- 4 "Partial Arrays" (I.e., Maps, Dictionaries, Etc.)
- Index Spaces
- 6 Larger Examples
- Conclusion

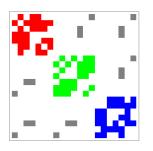
A Diagonal Block Array



```
\begin{array}{lll} \text{m1, m2, m3} & :: & \mathbb{Z}_7 \times \mathbb{Z}_7 \, \Rightarrow \, \mathbb{Z}_2 \\ \\ \text{dba} & :: & \mathbb{Z}_{21} \times \mathbb{Z}_{21} \, \Rightarrow \, \mathbb{Z}_2 \\ \\ \text{dba} & = & \\ & \text{unblock} \\ & \text{(fromTris} \\ & \text{( const[SET$^2(\mathbb{Z}_3)] (const[\mathbb{Z}_7 \times \mathbb{Z}_7] \ 0)} \\ & \text{, arr}[\mathbb{Z}_3] [\text{m1, m2, m3}] \\ & \text{, const[SET$^2(\mathbb{Z}_3)] (const[\mathbb{Z}_7 \times \mathbb{Z}_7] \ 0)} \\ & \text{)} \end{array}
```

• using \mathbb{Z}_2 to represent booleans (1 signifies edge between)

A Diagonal Block Array, with Sparse Off-Diagonals



```
m1, m2, m3 :: \mathbb{Z}_7 \times \mathbb{Z}_7 \Rightarrow \mathbb{Z}_2

dba' :: \mathbb{Z}_{21} \times \mathbb{Z}_{21} \Rightarrow \mathbb{Z}_2

dba' = unblock (arr [\mathbb{Z}_3 \times \mathbb{Z}_3]

[ m1 , \{..\} \underline{DFLT} 0, \{..\} \underline{DFLT} 0

, \{..\} \underline{DFLT} 0, \{..\} \underline{DFLT} 0, \{..\} \underline{DFLT} 0

, \{..\} \underline{DFLT} 0, \{..\} \underline{DFLT} 0, \{..\} \underline{MS}
```

 if adjacency matrix of graph, 3 graphs with sparse connectivity between

An Unfortunate Situation

- Many data-structures have "clean" decompositions
 - These support divide-and-conquer algorithms
 - E.g., lists (head & tail, split, ...), queues, binary-trees (left & right), etc.
- Graphs are not one of these
 - Though we do have some divide-and-conquer schemes, such as map-reduce.
- How might we decompose the graph just shown?
 - Has "obvious locality"

One Method to Divide-And-Conquer Graphs

Incomplete Graphs:

One Method to Divide-And-Conquer Graphs

Incomplete Graphs:

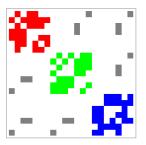
• ... generalize graphs

One Method to Divide-And-Conquer Graphs

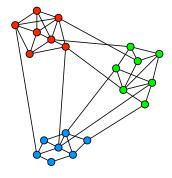
Incomplete Graphs:

- ... generalize graphs
- ... allow us to group vertices (similar to super-vertices)

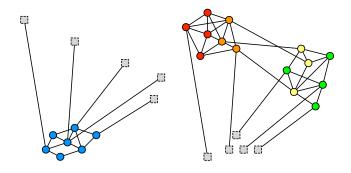
The adjacency matrix of graph:



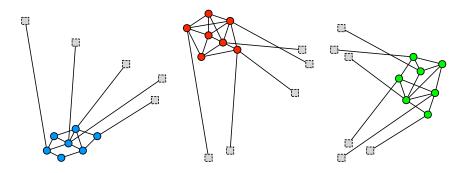
The graph:



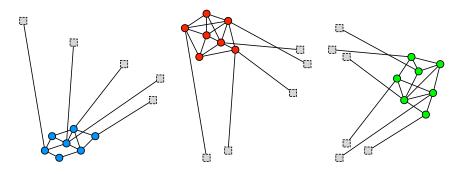
Extract blue nodes; edges get "split in half":



Extract green nodes, edges get "split in half":

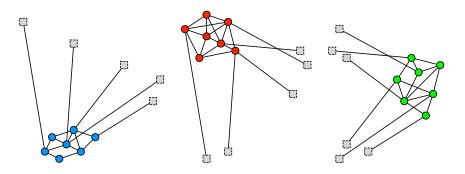


Extract green nodes, edges get "split in half":



• Can merge IGs in any order (associative and commutative)

Extract green nodes, edges get "split in half":



- Can merge IGs in any order (associative and commutative)
- Can do computations on IGs, computations updated when we merge

- Introduction
- Capturing Structure
- 3 Representations
- 4 "Partial Arrays" (I.e., Maps, Dictionaries, Etc.)
- Index Spaces
- 6 Larger Examples
- Conclusion

Observations

- Multiple features are working synergistically:
 - smart/const arrays (with a default tagged rep.)
 - bijections
 - expressive types
 - @, the expansion operator
- Novel features
 - powerful index space transformations
 - use of sums in an array language
 - type system (more expressive than most)
- Results
 - separation of interface and representation
 - expressive, high level array transformations
 - get views, and in-place updates, when we program with bijections

Conclusion

- In the paper, we discuss many more aspects of the language
 - types
 - functors (map-like functions)
 - program laws and transformations
 - the four combinatorial collections: SET, VEC, PERM, MSET
- Current status of project
 - A proof of concept embedded in Haskell
 - A few graph algorithms (Borůvka, Triangle Counting)
 - See the github project: https://github.com/GaloisInc/lpg
- Future
 - Need to "test drive" on more algorithms
 - Exploring yet more expressiveness in the type system
 - ...

Thank You

Supplementary Slides

sparse rep. of total arrays

```
{1: 2, 5: 3, 7: 0, ...} DFLT 0
                                                        :: \mathbb{Z}_{50} \Rightarrow \mathbb{Z} , ||=2 \times nn^2
```

sparse rep. of total arrays

```
\{1: 2, 5: 3, 7: 0, \ldots\} DFLT 0: \mathbb{Z}_{50} \Rightarrow \mathbb{Z}, ||=2 \times nn^2
```

```
t2p(@p2t{1: 2, 5: 3, 7: 0, ...}) :: \mathbb{Z}_{50} \Longrightarrow \mathbb{Z}, |=1+50|
```

sparse rep. of total arrays

```
\{1: 2, 5: 3, 7: 0, \ldots\} DFLT 0: \mathbb{Z}_{50} \Rightarrow \mathbb{Z}, ||=2 \times nn^2
```

```
t2p(@p2t{1: 2, 5: 3, 7: 0, ...}) :: \mathbb{Z}_{50} \Rightarrow \mathbb{Z}, ||=1+50
```

sparse rep. of total arrays

```
\{1: 2, 5: 3, 7: 0, \ldots\} DFLT 0: \mathbb{Z}_{50} \Rightarrow \mathbb{Z}, ||=2 \times nn^2
```

```
t2p(@p2t{1: 2, 5: 3, 7: 0, ...}) :: \mathbb{Z}_{50} \Longrightarrow \mathbb{Z}, ||=1+50
```

sparse rep. of total arrays

```
\{1: 2, 5: 3, 7: 0, \ldots\} DFLT 0: \mathbb{Z}_{50} \Rightarrow \mathbb{Z}, ||=2 \times nn^2
```

```
t2p(@p2t{1: 2, 5: 3, 7: 0, ...}) :: \mathbb{Z}_{50} \Longrightarrow \mathbb{Z}, ||=1+50|
```

What is a graph?

Graphs parameterized by vertex and edge:

```
type DG v e = VEC^2(v) \implies e -- directed
type UG v e = MSET^2(v) \implies e -- undirected
```

No edge data:

```
type DG' v = VEC<sup>2</sup>(v) \Rightarrow \mathbb{Z}_2 -- directed
type UG' v = MSET^2(v) \Rightarrow \mathbb{Z}_2 -- undirected
```

mapRes compare (smart[t] id)

```
mapRes compare (smart[t] id)
= smart[t] (compare . id)
```

```
mapRes compare (smart[t] id)
= smart[t] (compare . id)
= smart[t] compare
```

```
mapRes compare (smart[t] id)
= smart[t] (compare . id)
= smart[t] compare
= smart[t] (id . compare)
```

```
mapRes compare (smart[t] id)
= smart[t] (compare . id)
= smart[t] compare
= smart[t] (id . compare)
= mapIdx compare (smart[t] id)
```

```
compare :: VEC^2(a) \iff LT: SET^2(a) = COMPARE(a) + EQ: a + GT: SET^2(a)
```

```
compare :: VEC^2(a) \iff LT: SET^2(a) = COMPARE(a) + EQ: a + GT: SET^2(a)

less :: VEC^2(a) \iff LT: SET^2(a) + GTE: MSET^2(a)
```

```
compare :: VEC^2(a) \iff LT: SET^2(a) = COMPARE(a)
                             + EQ: a
                             + GT: SET^2(a)
less :: VEC^2(a) \iff LT: SET^2(a)
                            + GTE: MSET<sup>2</sup>(a)
         :: VEC^2(a) \iff
eq
                          EQ: a
                             + NE: PERM<sup>2</sup>(a)
```

```
compare :: VEC^2(a) \iff LT: SET^2(a) = COMPARE(a)
                                    + EQ: a
                                    + GT: SET^2(a)
less :: VEC^2(a) \iff LT: SET^2(a)
                                   + GTE: MSET<sup>2</sup>(a)
          :: VEC^2(a) \iff
eq
                                 EQ: a
                                    + NE: PERM<sup>2</sup>(a)
\mathbb{Z}(\mathsf{n}\times\mathsf{m}) \iff \mathbb{Z}_n \times \mathbb{Z}_m
```

```
smart[\mathbb{Z}_4] id :: \mathbb{Z}_4 \Rightarrow \mathbb{Z}_4
  (0,0) (0,1) (0,2) (0,3)
  (1,0) (1,1) (1,2) (1,3)
  (2,0) (2,1) (2,2) (2,3)
  (3,0) (3,1) (3,2) (3,3)
```

```
smart[\mathbb{Z}_4] id :: \mathbb{Z}_4 \Rightarrow \mathbb{Z}_4
   (0,0) (0,1) (0,2) (0,3)
   (1,0) (1,1) (1,2) (1,3)
   (2,0) (2,1) (2,2) (2,3)
   (3,0) (3,1) (3,2) (3,3)
mapRes compare' (smart[\mathbb{Z}_4] id) :: \mathbb{Z}_4 \Rightarrow \mathsf{LT} + \mathsf{EQ} + \mathsf{GT}
   EQ
                    LT
                             LT
           LT
   GT
           ΕQ
                    LT
                         LT
   GT
           GT
                EQ
                          LT
   GT
           GT
                    GT
                             EQ
```

```
mapRes compare' (smart[\mathbb{Z}_4] id) :: \mathbb{Z}_4 \Rightarrow LT + EQ + GT
  EQ
         LT
                LT
                       LT
  GT
         EΟ
            LT LT
  GT
         GT
            EO
                    LT
  GT
         GT
                GT
                       ΕO
mapRes compare (smart[\mathbb{Z}_4] id) :: \mathbb{Z}_4 \Rightarrow \mathsf{COMPARE}(\mathbb{Z}_4)
  EQ:0 LT:(0,1) LT:(0,2) LT:(0,3)
  GT:(0,1) EQ:1 LT:(1,2) LT:(1,3)
  GT:(0,2) GT:(1,2) EQ:2 LT:(2,3)
  GT:(0,3) GT:(1,3) GT:(2,3)
                                     EQ:3
```

```
mapRes compare (smart[\mathbb{Z}_4] id) :: \mathbb{Z}_4 \Rightarrow \mathsf{COMPARE}(\mathbb{Z}_4)

EQ:0 LT:(0,1) LT:(0,2) LT:(0,3)

GT:(0,1) EQ:1 LT:(1,2) LT:(1,3)

GT:(0,2) GT:(1,2) EQ:2 LT:(2,3)

GT:(0,3) GT:(1,3) GT:(2,3) EQ:3
```