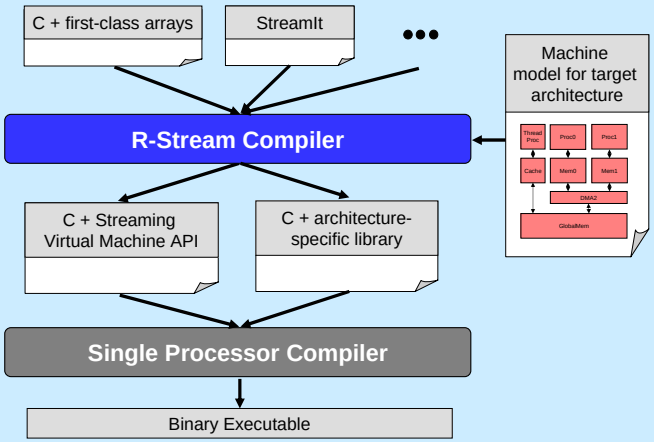
R-Stream: Compiler Technology for Next Generation HPEC

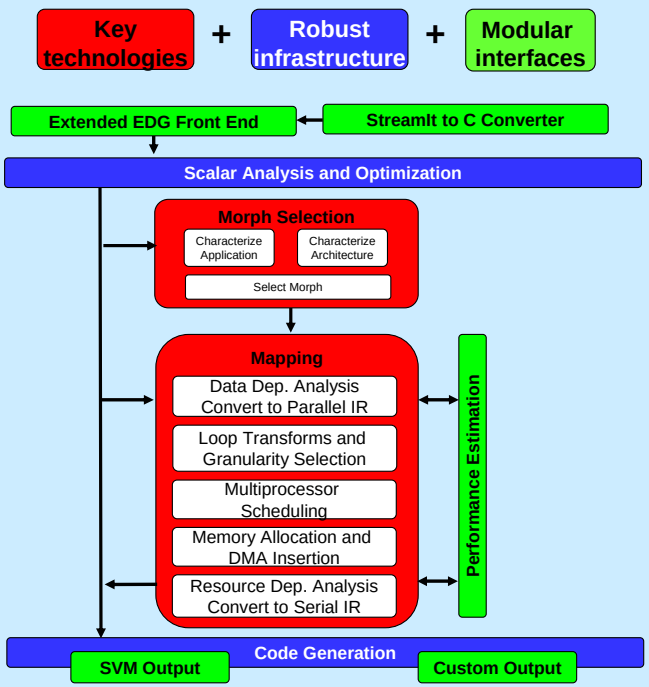
Reservoir Labs Inc.

Role in Tool Chain

R-Stream is a source-to-source compiler intended to **augment** an **existing** single processor **tool chain**.



Compiler Structure



Compiler Tech. for HPEC

R-Stream compiler technology **automatically maps** applications to HPEC architectures with:

- **Multiple processor** cores
- Distributed on-chip memories w/ **DMA**
- **Reconfigurable** processors and memories

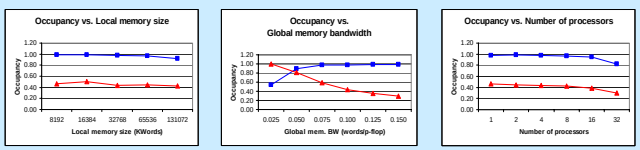
R-Stream **optimizes the whole application**, e.g. reducing memory traffic between kernels, unlike using a library alone.

R-Stream maps one C program to multiple targets, for **faster, cheaper, more reliable development** than mapping by hand.

Early Results

Early results show **efficient mappings** over a wide range of architectural parameters:

	Ex. Shown	TRIPS	Smart Mem.	RAW	M-Chip (Not actual)	Imag-ine
Stream Processors	4	4	4	16	8	8
FP ALUs	8	16	2	1	8	6
Frequency	500	1000	500	420	1000	250
Gflops	16.0	64.0	4.0	6.7	64.0	12.0
Local Memory Size (words)	32768	65536	24576	8192	512 (n per proc)	64000
Global Memory BW (bytes/ns)	1.6	0.262	4	1	4	2.3
Global Memory BW (words/p-flop)	0.100	0.001	0.250	0.037	0.016	0.048

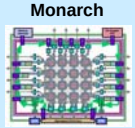


Supports Diverse Architectures

R-Stream prototype supports a **large class of architectures** via a **flexible machine model**, including:

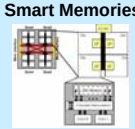
MIT RAW

ISI / Raytheon Monarch



UT Austin TRIPS

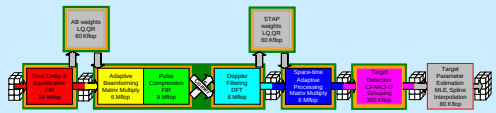
Stanford Smart Memories



Prototype 2.0 Mapper

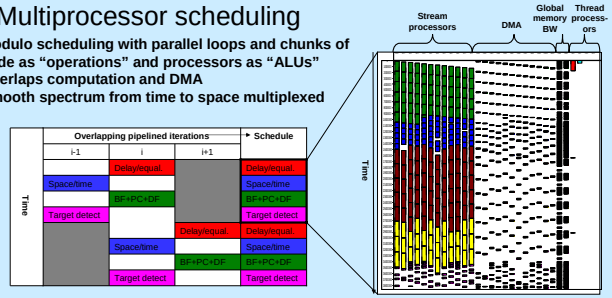
1. Transform loops for locality, determine granularity

- Goal is maximizing data that can live in local memory or local memories
- Interchange and partially fuse parallel outer loops
- Classify communications as local memory, inter-processor, or global memory
- Single-processor grains contain local memory communication
- Multi-processor grains contain communication between local memories



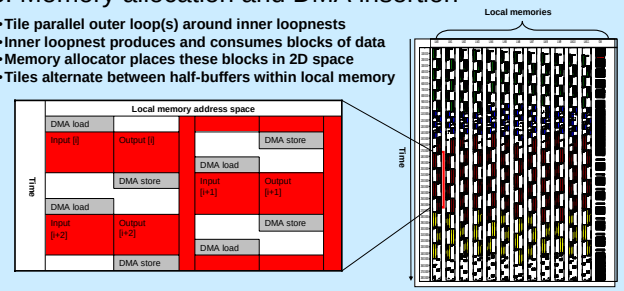
2. Multiprocessor scheduling

- Modulo scheduling with parallel loops and chunks of code as "operations" and processors as "ALUs"
- Overlaps computation and DMA
- Smooth spectrum from time to space multiplexed



3. Memory allocation and DMA insertion

- Tile parallel outer loop(s) around inner loopnests
- Inner loopnest produces and consumes blocks of data
- Memory allocator places these blocks in 2D space
- Tiles alternate between half-buffers within local memory



Innovative 3.0 Technology

R-Stream prototype 3.0, currently in development, will produce even **more efficient mappings** for a **wider range of applications** by leveraging:

- SRE-based internal representation to **eliminate false dependences**
- Affine partitioning framework to discover **maximum degrees of parallelism** in application
- Unified/constraint-based mapping to **avoid phase-ordering**.