A simple 1-D multiply and add kernel

[16]:
```python
#This is to check if fma operations are used instead of typical multiply and then add

@cuda.jit
def multiply_by_scalar_and_add(arr, scalar, result):
    idx = cuda.grid(1)

    if idx < arr.size:
        result[idx] = (arr[idx] * scalar) + scalar
```

[17]:
```python
def test_multiply_scalar_and_add():
    arr = np.random.random(10000000)
    scalar = 8.8
    result = np.zeros_like(arr)

    tpb = 1024
    d_result = cuda.to_device(result)
    multiply_by_scalar_and_add[return_blocks_and_threads(tpb, arr.size), tpb](
        cuda.to_device(arr),scalar, d_result)

    print(d_result.copy_to_host())
```

[18]:
```python
test_multiply_scalar_and_add()
```

```
[14.19856061 11.04312628 11.38904771 ...  9.80525965 17.13052089
 12.55831722]
```

```
//
// Generated by NVIDIA NVVM Compiler
//
// Compiler Build ID: CL-34097967
// Cuda compilation tools, release 12.4, V12.4.131
// Based on NVVM 7.0.1
//

.version 8.4
.target sm_61
.address_size 64

      // .globl
_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01Ew
1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arra
yIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE
.visible .global .align 4 .u32
_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01Ew
1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arra
yIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE__errcode__;
```

```
.visible .global .align 4 .u32
_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01Ew
1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arra
yIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE__tidx__;
.visible .global .align 4 .u32
_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01Ew
1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arra
yIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE__ctaidx__;
.visible .global .align 4 .u32
_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01Ew
1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arra
yIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE__tidy__;
.visible .global .align 4 .u32
_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01Ew
1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arra
yIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE__ctaidy__;
.visible .global .align 4 .u32
_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01Ew
1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arra
yIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE__tidz__;
.visible .global .align 4 .u32
_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01Ew
1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arra
yIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE__ctaidz__;
.common .global .align 8 .u64
_ZN08NumbaEnv8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw0
1Ew1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5A
rrayIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE;

.visible .entry
_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01Ew
1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arra
yIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE(
        .param .u64
_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01Ew
1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arra
yIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE_param_0,
        .param .u64
_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01Ew
1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arra
yIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE_param_1,
        .param .u64
_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01Ew
1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arra
yIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE_param_2,
        .param .u64
_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01Ew
1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arra
yIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE_param_3,
```

```
	.param .u64
_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01Ew
1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arra
yIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE_param_4,
	.param .u64
_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01Ew
1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arra
yIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE_param_5,
	.param .u64
_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01Ew
1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arra
yIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE_param_6,
	.param .f64
_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01Ew
1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arra
yIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE_param_7,
	.param .u64
_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01Ew
1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arra
yIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE_param_8,
	.param .u64
_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01Ew
1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arra
yIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE_param_9,
	.param .u64
_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01Ew
1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arra
yIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE_param_10,
	.param .u64
_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01Ew
1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arra
yIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE_param_11,
	.param .u64
_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01Ew
1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arra
yIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE_param_12,
	.param .u64
_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01Ew
1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arra
yIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE_param_13,
	.param .u64
_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01Ew
1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arra
yIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE_param_14
)
{
	.reg .pred  %p<2>;
	.reg .b32   %r<4>;
	.reg .f64   %fd<4>;
```

```
	.reg .b64 	%rd<20>;


	ld.param.u64 	%rd6,
[_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01E
w1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arr
ayIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE_param_2];
	ld.param.u64 	%rd2,
[_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01E
w1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arr
ayIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE_param_4];
	ld.param.u64 	%rd3,
[_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01E
w1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arr
ayIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE_param_5];
	ld.param.f64 	%fd1,
[_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01E
w1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arr
ayIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE_param_7];
	ld.param.u64 	%rd4,
[_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01E
w1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arr
ayIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE_param_12];
	ld.param.u64 	%rd5,
[_ZN6cudapy8__main__26multiply_by_scalar_and_addB2v5B96cw51cXTLSUwv1sCUt9Uw01E
w1NRRQPKiLTj0gIGIFp_2b2oLQFEYYkHSQB1OQAk0Bynm21OizQ1K0UoIGvDpQE8oxrNQE_3dE5Arr
ayIdLi1E1C7mutable7alignedEd5ArrayIdLi1E1C7mutable7alignedE_param_13];
	mov.u32 	%r1, %tid.x;
	cvt.s64.s32 	%rd7, %r1;
	mov.u32 	%r2, %ntid.x;
	mov.u32 	%r3, %ctaid.x;
	mul.wide.s32 	%rd8, %r2, %r3;
	add.s64 	%rd1, %rd8, %rd7;
	setp.ge.s64 	%p1, %rd1, %rd6;
	@%p1 bra 	$L__BB0_2;

	cvta.to.global.u64 	%rd9, %rd2;
	shr.s64 	%rd10, %rd1, 63;
	and.b64 	%rd11, %rd10, %rd3;
	add.s64 	%rd12, %rd11, %rd1;
	shl.b64 	%rd13, %rd12, 3;
	add.s64 	%rd14, %rd9, %rd13;
	ld.global.f64 	%fd2, [%rd14];
	fma.rn.f64 	%fd3, %fd2, %fd1, %fd1;
	and.b64 	%rd15, %rd10, %rd5;
	add.s64 	%rd16, %rd15, %rd1;
	cvta.to.global.u64 	%rd17, %rd4;
	shl.b64 	%rd18, %rd16, 3;
	add.s64 	%rd19, %rd17, %rd18;
```

```
        st.global.f64     [%rd19], %fd3;

$L__BB0_2:
        ret;

}
```

Key insights:-

1) No shared memory - Doesn't make sense given the input is 1D array
2) You can see a fp64 ops. Which is an opportunity to fine tune to fp32
3) Global memory access all along
4)
```
        mov.u32     %r1, %tid.x;
        cvt.s64.s32 %rd7, %r1;
        mov.u32     %r2, %ntid.x;
        mov.u32     %r3, %ctaid.x;
        mul.wide.s32      %rd8, %r2, %r3;
        add.s64     %rd1, %rd8, %rd7;
        setp.ge.s64 %p1, %rd1, %rd6;
        @%p1 bra    $L__BB0_2;
```
This is all equivalent to cuda.grid() and your bound check. Honestly respect the simplicity for programmers. Again not much to do here.

5) **Most important in my opinion - The use of FMA at** `fma.rn.f64      %fd3, %fd2, %fd1, %fd1;`

FMA is Fused multiply add.