Radio-Frequency Matrix-Vector Multiplication

6.2300 Class Project: Detailed Plan
TeamID: RFCompute

Dominic Stewart-Guido
EECS
dominics@mit.edu

Ishaan Vohra
EECS
ivohra@mit.edu

Fadi Farag
EECS
fadif@mit.edu

***Abstract*—As the world continues to use radio-frequency-based devices, such as GPS satellites, MRI machines, speakers, and radar surveillance systems, it becomes increasingly important to be able to perform quick computations for practice and research. In an effort to increase the throughput of these systems, we propose a novel approach to conducting vector-matrix multiplication on a radio-frequency level with a wide nearly instantaneous band. An attenuator scales an incoming voltage wave by some transmission coefficient. Two attenuators in series each attenuate by their own coefficient, therefore creating a total transmission coefficient equal to the product of the two individual coefficients. This concept, in combination with Wilkinson power splitters and Wilkinson power combiners, processes computations in terms of the transmission coefficient.**

I. Introduction & Motivations

1) Our project uses electromagnetic waves to perform a matrix-vector multiplication. It is thus a proof of concept for RF analog computation. Analog computers are superior to digital computers for many applications. They render igital-to-analog conversion unnecessary and avoid issues of sampling and aliasing, which compromise information [1], are much quicker, as they operate in real time, and can often save significant energy compared to digital computers by performing operations passively using physical phenomena instead of the energy-intensive switching of transistor states. They are also excellent for complex simulations, as useful mathematical relationships such as differential equations and linearity (superposition) can be built into their physics [2], [4]. As we progress towards more complex computational problems, incorporating analog or mixed computation has the potential to greatly increase efficiency, speed, and accuracy for certain applications, as exemplified in Fig. 1. Our project is an important first step towards this goal, as matrix operations are crucial for many computations.
2) This project is not the first effort to address computation times through hardware means. For instance, Bandyopadhyay et. al. [10] use integrated photonics to create coherent matrix multiplication units that process matrix-vector multiplications for deep neural networks. The matrix multiplication unit's interference in a Mach–Zehnder interferometer mesh allow them to avoid off-chip conversions and accelerate performance.

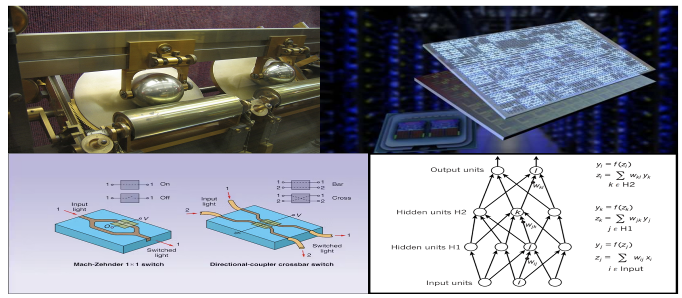

Fig. 1. Top left: a differential analyzer – a 20th-century analog computer which could solve certain differential equations. Top right: an artistic rendition of a future large-scale optical computer performing highly-parallelized calculations. Bottom left: a Mach-Zehnder interferometer – a common computational component used in optical analog computers. Bottom right: a diagram of a deep neural network – a major application of highly-parallelized linear computing, which optical analog computers excel at. [2]–[5]

Paliy et. al. address computation time through analog means [11]. This group developed an analog vector-matrix multiplier using current mirrors with adjustable current output to current input ratios as representations of weight multiplications.
3) The heart of our computational mechanism –variable attenuators – have also been employed by other independent research teams for different applications, such as precise beamforming and transmitarray systems. The first group, led by Tie Jun Cui at Southeast University (described in Li et. al [12]), realized a C-band programmable transmitarray using PIN diode variable attenuators for beamforming and power-allocation. The second group, from CNR-IMM in Italy, developed RF-MEMS variable attenuators for beyond-5G beamforming (described in Tagliapietra et. al. [13]), making use of 3-bit attenuators to fine tune the fixed attenuations of pi-shaped resistors.
4) To carry out matrix-vector multiplication in $\mathbb{R}^2$, we separate the process into computing two dot products, making use of multiple variable attenuators to perform multiplication and a Wilkinson power combiner to perform scalar addition. Each variable attenuator multiplies an ingoing sinusoidal voltage signal by a tunable attenuation; after two voltage signals pass through two

attenuators each, they are combined to sum their voltage amplitudes together (see Section II for more details). Results can be found in Section IV.

II. Approach (aka Materials & Methods)

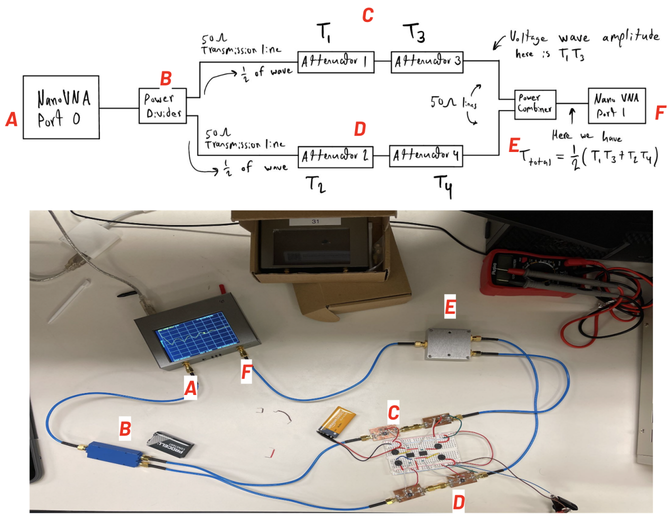

Fig. 2. Block Diagram and physical setup of Apparatus. Given $\begin{pmatrix} c & d \\ e & f \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} ac+bd \\ ae+ef \end{pmatrix}$, setting $T_1 = a$, $T_2 = b$, $T_3 = c$, and $T_4 = d$ computes the first entry of the matrix-vector product, $ac + bd$. [7]

This being a proof of concept, we deemed computing one dot product to be sufficient for this project, noting that any matrix multiplication could be calculated by expanding our apparatus. (For example, by inserting Wilkinson dividers after T_1, T_2.) An attenuator scales an incoming voltage wave by some transmission coefficient T_i. If two attenuators with transmission coefficients T_1, T_2 are cascaded, an incoming wave is scaled fist by T_1, then by T_2, creating a total transmission coefficient of $T_1 \times T_2$ at their output. In this way, multiplication is implemented through cascaded attenuators. Variable attenuators allow us to change each T_i to tune the numbers we are multiplying. Although $T_i \leq 1$, any numbers can implemented if we divide and multiply them all by a normalization constant before and after the calculation. Although we did not discuss transmission much in 6.2300, we extensively covered its cousin, the reflection coefficient.

We used the AT-110TR voltage variable attenuator, which can be purchased online. For proper control of and connection to each chip, we milled PCBs containing connections for the chip's ground, power, and control pins, and 50Ω coplanar waveguides for the chip's RF pins. Each waveguide was terminated on one end by an SMA connector and on the other by a 1nF SMD DC blocking capacitor. We soldered all components onto our PCB by hand, as well as wires for DC connections. A breadboard circuit was used to implement 4 potentiometer voltage dividers, which we buffered and connected to the chip's control pins, allowing us to accurately tune each T_i.

To measure transmission, we use a NanoVNA, obtained from our 6.2300 instructors. Port 0 of the NanoVNA is connected to a power divider, which splits its sensing wave along two paths, as seen in the diagram. Splitting the wave in half contributes a factor of $\frac{1}{2}$ to each path. The paths each have a pair of attenuators, which compute the products ac, db. Their outputs are then combined using a power combiner, which, given that the two waves are in phase, will add their amplitudes, producing a total transmission coefficient of $T_{total} = \frac{1}{2}(ac+bd)$ for the network, finishing the computation. This T_{total} is measured by connecting the combiner's output to port 1 of the NanoVNA. The NanoVNA should have its thru calibration performed on the setup with $T_i = 1$ for each attenuator so that the system's internal losses are not included in its calculations.

To ensure the signals added in phase, and to mitigate unwanted reflections and losses, all connections between RF elements were made using identical 50Ω coaxial cables and SMA connectors. This process of matching transmission lines was a significant portion of 6.2300, although we implemented coplanar transmission lines on these PCBs, as opposed to the microstrip lines we usually fabricated in our labs. A Wilkinson power splitter can act as both a divider and a combiner, so our setup used two Wilkinson splitters, which we obtained from our TA Réouven Assouly.

III. Legality, Safety and Ethics

Our NanoVNA sweeps over 100kHz to 1.5GHz, but attenuates at 900MHz. Since the NanoVNA already operates at power levels below FCC MPE limits and our attenuator can only decrease the output power, it it safe to assume that the NanoVNA will remain FCC-compliant. To verify this, we will look at Table 1 in FCC Title 47 CFR § 1.1310(e)(1) which limits the power density of any emissions, depending on their frequency range. At an operating frequency of 900MHz, our power density must be less than 3 $\mathrm{mW/cm^2}$ or even 1 $\mathrm{mW/cm^2}$ between 30MHz and 300MHz for controlled exposure. This becomes 0.6 $\mathrm{mW/cm^2}$ or even 0.2 $\mathrm{mW/cm^2}$, respectively, for public exposure. Our NanoVNA's power output is well below this limit, and each variable attenuator can further reduce the output by orders of magnitude, as shown in Figure 3. To further mitigate radiation exposure risk, we plan to maintain a few centimeters of clearance from any active RF port. Additionally, under FCC Title 47 §15.205, our frequency of 900MHz is permitted and 60MHz away from the closest restricted band. Regarding broader stakeholder impacts, direct stakeholders include lab personnel and guest researchers unfamiliar with RF safety, and indirect stakeholders include adjacent lab occupants whose sensitive instruments could be affected by unlikely RF leakage. To mitigate these risks, we will solicit feedback from stakeholders before initial operation to address concerns, and ensure they are notified about the risks of our experiment. We will proceed with our experiment only if their consent is given.

IV. Results

For analysis of Table 1 and Figure 4, see Discussion section. The attenuation-over-frequency plot shown in Figure 3 is meant to hint at the potential expansion of this matrix-vector multiplier via frequency multiplexing. Though our matrix-vector multiplier was designed to operate at 900 MHz according to the optimal performance curve of our attenuators (and demonstrates roughly constant attenuation over frequency, as shown in Figure 3), a more robust attenuation device with control over frequency-dependent attenuation could enable much more powerful computation. Due to the linearity inherent in Maxwell's equations, dot products with frequency-controllable attenuators would enable many more calculations to be performed in parallel, with the exact number depending on the bandwidth of each frequency band. This immense parallelization could truly have a large impact on signal processing applications which require many matrix operations to be performed very rapidly.

a	c	b	d	T_{total}	Measured Result	Absolute Error Magnitude
1	1	1	1	1.00	1.00	0.00
1	1	0	1	0.50	0.51	0.01
1	1	0.32	1	0.66	0.69	0.03
1	1	0.14	1	0.57	0.58	0.01
0.32	0.32	0.32	0.32	0.10	0.10	0.00
0.14	0.14	0.14	0.14	0.02	0.04	0.02
Avg Error						0.01

TABLE I
SEVERAL SAMPLE DOT PRODUCT CALCULATION RESULTS FOR $T_{total} = \frac{1}{2}(ac + bd)$, DEMONSTRATING CORE FUNCTIONALITY.

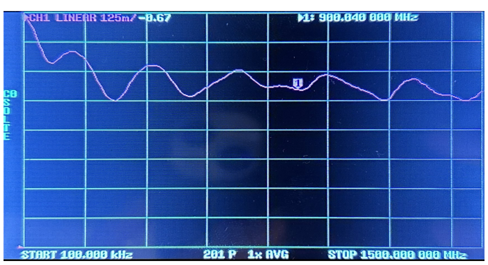

Fig. 3. Attenuation plot on NanoVNA showing attenuation (linear scale) over frequency, hinting at the ability of a larger and more powerful system to perform multiple calculations at once, multiplexing over frequency. [8]

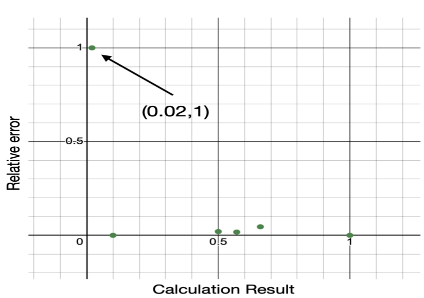

Fig. 4. Scatter plot of relative error of calculations performed in Table 1, demonstrating lower performance for smaller encoded numbers. [9]

V. DISCUSSION

Table 1 displays sample calculations of dot products performed with the matrix-vector multiplier. We observe very low absolute error (about 2 orders of magnitude less than the maximum value of encoded numbers), thereby demonstrating the core functionality of the system as an effective method of performing computations.

The main limitation of our matrix-vector multiplier, as shown in Figure 4, is the relative error of the calculations. By simply recasting the absolute errors calculation in Table 1, it can be shown that - as is commonly an issue for many analog computation devices – performing calculations with small numbers can lead to especially large relative error. In our case, this is caused by both the noise background, as well as the maximum attenuation limit of our attenuators. In reality, when our attenuators are set to encode a numerical value of 0, they are really attenuating by about -40 dB, which corresponds to an encoded value of about 0.01, not 0. This offset is the main cause of our errors as seen in Table 1, and also causes the relative error for calculations with small values (as depicted by the arrow in Figure 4) to be much larger, often unacceptably inaccurate.

It is likely that attenuators with a greater maximum attenuation would help to mitigate this issue, and enable encodings of numbers closer to true 0, and therefore lower relative error for small values. Nevertheless, the issue of larger relative errors for small number encodings remains an issue that plagues many analog computer designs.

VI. CONCLUSIONS

Setting $T_i = 0.32$ for each attenuator results in the following calculation (when rounded to two decimal places, the NanoVNA's precision): $\frac{1}{2}(0.32 \times 0.32 + 0.32 \times 0.32 = 0.01$, which it performs exactly as desired. Our average error across all the calculations we did in our dot product table was only 0.01, suggesting that it is highly accurate. This proof of concept device sets the stage for more complex RF matrix-vector multipliers, which could further improve performance through frequency multiplexing. Transmission coefficients are ubiquitous in wave phenomena, so there are countless ways to implement this computation other than through potentiometers.

REFERENCES

[1] J. Schesser (2018). NJIT BME310 Biomedical Computing, Lecture 6: Sampling and Aliasing [pdf slides] https://web.njit.edu/~joelsd/Fundamentals/coursework/BME310computingcw6.pdf
[2] A. Dingley. "Differential Analyzer." Wikipedia. Accessed: April 22, 2025. https://en.wikipedia.org/wiki/Differential_analyser
[3] M. Heck. "Optical computers light up the horizon." Phys.org. Accessed: April 22, 2025. https://phys.org/news/2018-03-optical-horizon.html
[4] T. Higgins. "Computing with photons at the speed of light." Laser Focus World. Accessed: April 22, 2025. https://www.laserfocusworld.com/optics/article/16553319/computing-with-photons-at-the-speed-of-light
[5] Yann Lecun, Yoshua Bengio, Geoffrey Hinton. Deep learning. Nature, 2015, 521 (7553), pp.436-444. ff10.1038/nature14539. hal-04206682
[6] Author: Office of Engineering and TechnologyFederal Communications Commission, "Understanding the FCC Regulations Low-Power, Non-Licensed Transmitters", OET BULLETIN NO. 63October 1993. https://transition.fcc.gov/bureaus/oet/info/documents/bulletins/oet63/oet63rev.pdf

[7] F. Farag, D. Stewart-Guido, I. Vohra, “RF Matrix Vector Multiplier Apparatus.”
[8] F. Farag, D. Stewart-Guido, I. Vohra, “Attenuation over frequency plot.”
[9] F. Farag, D. Stewart-Guido, I. Vohra, “Relative error scatter plot.”
[10] S. Bandyopadhyay, A. Sludds, S. Krastanov, et al. Single-chip photonic deep neural network with forward-only training. Nat. Photon. 18, 1335–1343 (2024). https://doi.org/10.1038/s41566-024-01567-z
[11] M. Paliy, S. Strangio, P. Ruiu, T. Rizzo and G. Iannaccone, ”Analog Vector-Matrix Multiplier Based on Programmable Current Mirrors for Neural Network Integrated Circuits,” in IEEE Access, vol. 8, pp. 203525-203537, 2020, doi: 10.1109/ACCESS.2020.3037017.
[12] L., He, et al. ”Design of programmable transmitarray antenna with independent controls of transmission amplitude and phase.” IEEE Transactions on Antennas and Propagation 70.9 (2022): 8086-8099.
[13] G. Tagliapietra et al. Discussion and Demonstration of RF-MEMS Attenuators Design Concepts and Modules for Advanced Beamforming in the Beyond-5G and 6G Scenario—Part 2. Micromachines, 15(7), 895. (2024): https://doi.org/10.3390/mi15070895