📄 Chapter 5.3: Capturing, Normalizing, and Preparing IQ Data from RTL-SDR

1. Purpose

When working with RTL-SDR or any SDR receiver, the received IQ samples are usually saved in a **raw binary file**. This file must be correctly **read, interpreted, normalized, reshaped, and corrected** before further DSP or research work can be performed.

The main goals are:

1. Correctly extract I (in-phase) and Q (quadrature) components.
2. Normalize samples into a meaningful numeric range (typically `-1` to `+1`).
3. Handle **hardware DC offset** or imbalance.
4. Ensure reshaping produces clean IQ pairs for FFTs, constellation diagrams, and higher-level processing.

---

2. Data Format (Why 127.5 Appears Everywhere)

- RTL-SDR saves IQ samples as **unsigned 8-bit integers** (`uint8`).
- Values range from **0** → **255**.
- But real signals must be **zero-centered** (negative and positive swings).
- To convert:

$$\text{normalized\_sample} = \frac{(\text{raw\_sample} - 127.5)}{127.5}$$

- Subtracting **127.5** recenters values around 0.
- Dividing by **127.5** scales values into `[-1, +1]`.
- Example:
  - Raw = 0 → (0 - 127.5) / 127.5 = -1

- Raw = 127.5 → 0
- Raw = 255 → +1

So after this, the samples are in proper floating-point range.

3. Two Levels of Correction

Here's the **critical clarification** about why sometimes we see corrections done **twice** — once on `raw`, and again on `I` and `Q`.

Level 1: Hardware Normalization (raw level)

- Removes unsigned storage bias (0–255 → -1..+1).
- Ensures signal swings are mathematically valid.
- Equation: `(raw - 127.5)/127.5`.

Level 2: Channel Balancing (I/Q level)

- Even after normalization, **hardware imperfections** remain:
  - DC offset (mean ≠ 0).
  - I/Q imbalance (different mean or gain between I and Q).
- So, we compute:

$$I_0 = I - \text{mean}(I)$$

$$Q_0 = Q - \text{mean}(Q)$$

This ensures both I and Q are strictly zero-centered.

🔑 Answer to your concern:
If we only subtract the global mean at the raw stage (`raw0 = raw - raw.mean()`), we do not account for I and Q differences individually. Thus, both levels are required for precision SDR research work.

4. Experiment: Using `raw - raw.mean()`

Suppose instead of `(raw - 127.5)/127.5`, we try:
python
raw0 = raw - raw.mean()

- This removes **global bias** across the entire dataset.
- But SDR data is structured as alternating I, Q samples.
- If I has mean = 0.01 and Q has mean = -0.02, their biases will **cancel** in the global mean.
- Net result: residual offsets in I/Q that distort constellation and FFT plots.

☞ That's why raw.mean() subtraction **is not enough** for serious SDR work.

5. Step-by-Step Python Code

Below is the **final combined code** with all stages clearly separated and documented.

python
import numpy as np
import matplotlib.pyplot as plt

1. Load raw RTL-SDR data

filename = r"C:\IQ_Data\Prac_1.bin"
fs = 2400000 # Sample rate in Hz
raw = np.fromfile(filename, dtype=np.uint8)
print("Length of raw data:", len(raw))

2. Convert to float and normalize to [-1, +1]

```
raw = raw.astype(np.float32)
raw = (raw - 127.5) / 127.5
print("Raw normalized:", raw[:10]) # Show first 10 samples
```

3. Optional global DC removal (rarely enough by itself)

```
raw0 = raw - raw.mean()
print("Raw0 global corrected, mean:", raw0.mean())
```

4. Ensure even length for IQ pairs

```
if len(raw) % 2 != 0:
raw = raw[:-1]
```

5. Reshape into I/Q pairs

```
IQ = raw.reshape(-1, 2)
I = IQ[:, 0]
Q = IQ[:, 1]
```

6. Channel-specific DC removal

```
I0 = I - I.mean()
Q0 = Q - Q.mean()
```

7. Diagnostics

```
print(f"Raw min/max after normalize: {raw.min():.3f}, {raw.max():.3f}")
print(f"Mean(I), Mean(Q): {I.mean():.4f}, {Q.mean():.4f} (before DC removal)")
```

print(f"Mean(I0), Mean(Q0): {I0.mean():.4f}, {Q0.mean():.4f} (after DC removal)")

8. Quick plot check

plt.figure(figsize=(6,6))
plt.scatter(I0[:5000], Q0[:5000], s=1, alpha=0.5)
plt.title("Constellation Plot (DC Removed)")
plt.xlabel("I")
plt.ylabel("Q")
plt.grid(True)
plt.show()

6. Typical Output

Length of raw data: 14516224
Raw normalized: [-0.992, -0.976, -0.960, ...]
Raw0 global corrected, mean: ~0.000
Raw min/max after normalize: -1.000, 1.000
Mean(I), Mean(Q): 0.0123, -0.0087 (before DC removal)
Mean(I0), Mean(Q0): 0.0000, 0.0000 (after DC removal)
Constellation plot: shows **symmetric scatter around zero** after proper correction.

7. Key Takeaways

1. **Normalization with 127.5** is required because RTL-SDR stores unsigned bytes.
2. Subtracting only `raw.mean()` is insufficient — I and Q need **separate correction**.
3. Double-handling (raw level + I/Q level) is **not redundancy**; it's best practice.
4. Always check with **plots** (FFT, constellation) to confirm correction worked.
5. This pipeline is the foundation for **further DSP research** (demodulation, drone-signal detection, etc.).