Selected Topics in Electrical Engineering: Flow Cytometry Data Analysis

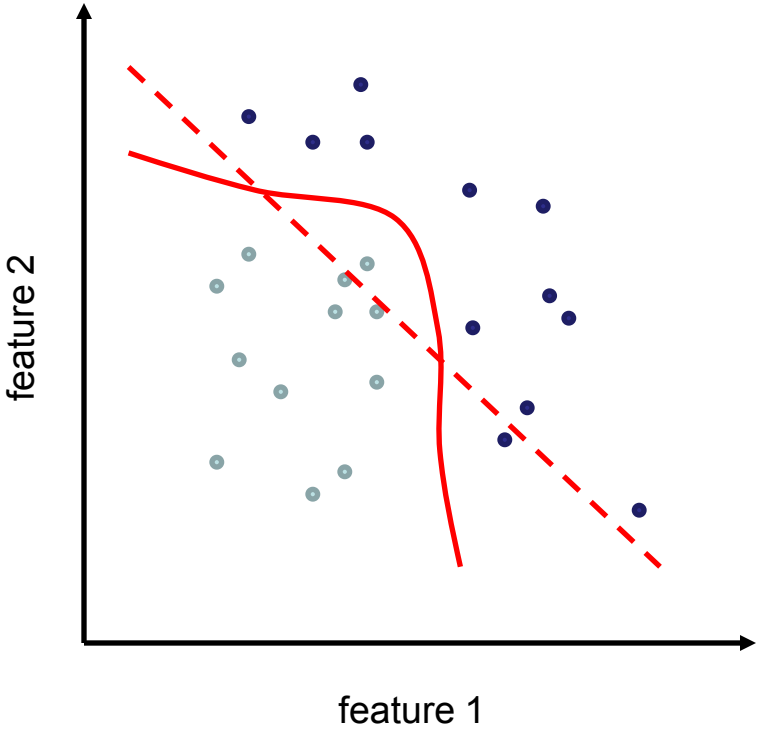
Bilge Karaçalı, PhD
Department of Electrical and Electronics Engineering
Izmir Institute of Technology

Outline

- Compensation and gating
  - Compensation
    - Spectral overlap
    - Two-color compensation
    - Compensation for three or more colors
  - Gating

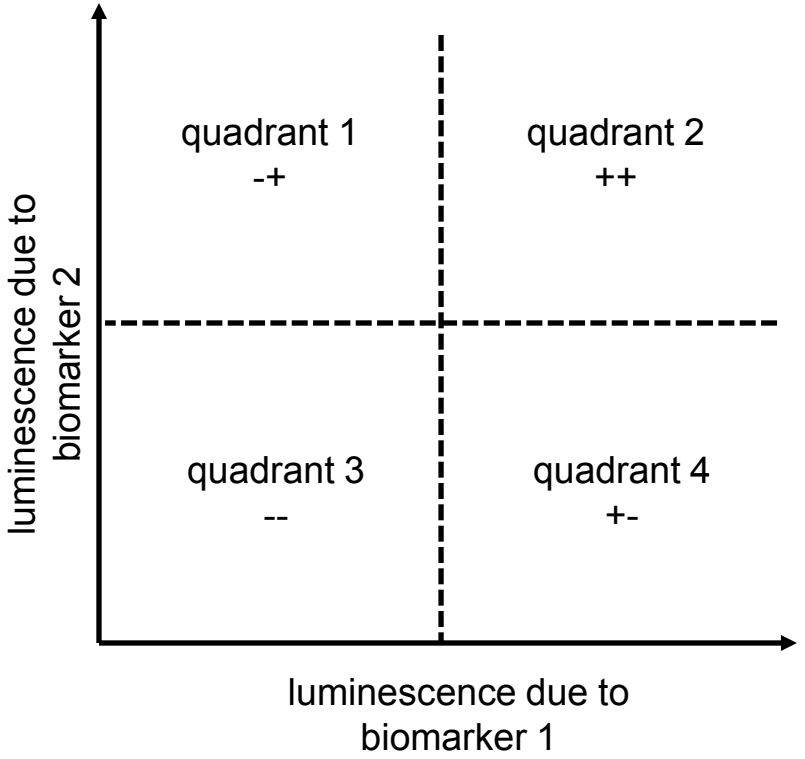
Feature Correlation in Pattern Recognition

- Correlation between the different features is useful for pattern recognition
  - Simultaneous variations between the feature vectors of the different classes reveal the separation boundaries
  - In effect, the separation boundaries locally encode the direction in the feature space that would transform the vectors of one class into the other



Feature Correlation in Flow Cytometry Data

- In flow cytometry applications;
  - The objective is to identify the cells that possess a biomarker from the other ones
    - CD4+ lymphocytes from the CD4- lymphocytes in AIDS
  - This is achieved easily (manually) if/when the +/- cell distributions appear in different quadrants
    - in the display of one fluochrome intensity against another
    - through the specific staining of the biomarkers by the fluochromes

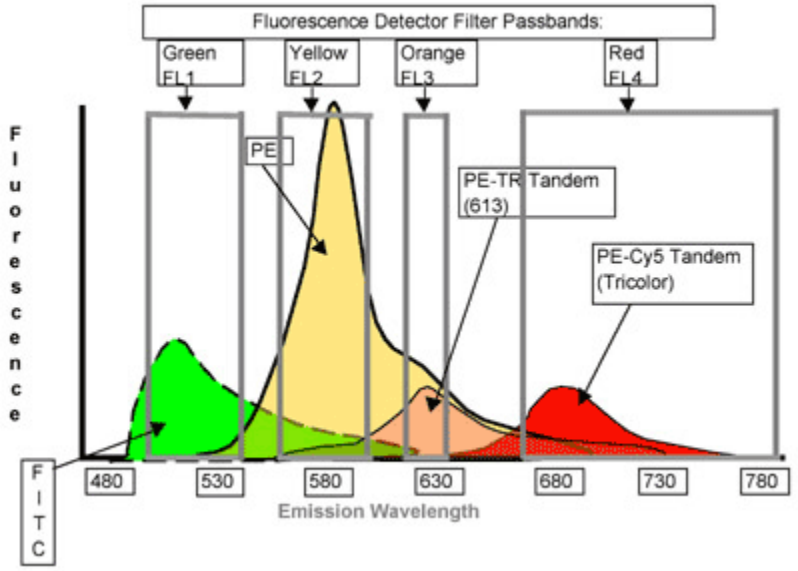


Feature Correlation in Flow Cytometry Data

- This means:
  - The (positive) correlation between the two luminescence measurements will
    - shift the +- cells up towards the ++ quadrant
    - shift the -+ cells to the right towards the ++ quadrant
  - The dispersion of the +-, -+ as well as ++ cells will be skewed
    - into a long, thin and pointy arc instead of a nice and round (Gaussian) distribution
    - due to the logarithmic scaling of the intensity measurements in the display
- The correlation between the intensity measurements needs thus to be removed from the data before further analysis

Feature Correlation in Flow Cytometry Data

- The correlation between the flurescence intensity measurements is caused by the **spectral overlap** between the emission spectra of the different fluorochromes
  - Fluorochrome presence is detected through optical filters centered at the peak of their emission spectra
  - Fluorochromes other than the intended one also contribute to the detected fluorescence
    - through long tails of their emission spectra that leaks into the corresponding optical bandpass filter
  - This causes increased fluorescence readings at the fluorescence detector
  - And the presence of the leaking fluorochrome makes the target fluorochrome appear in greater quantities than it is in actuality

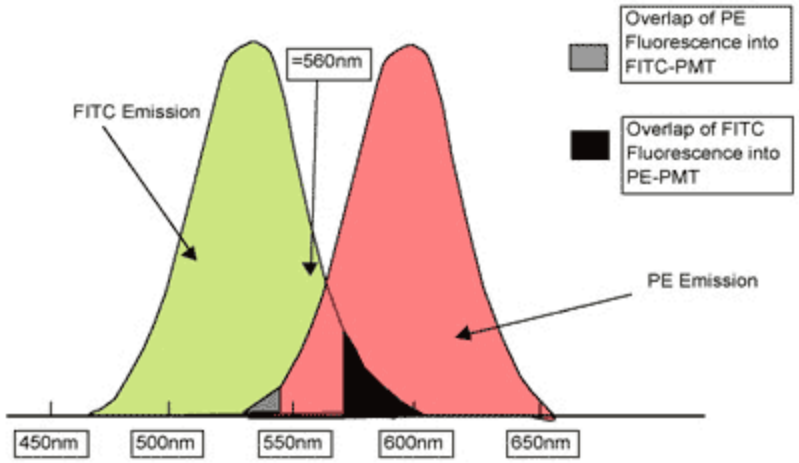


Source: http://crl.berkeley.edu/compensation.html

Spectral Overlap

- Technically,

  Spectral overlap is the phenomenon by which the emission spectra of the different fluorochromes intersect over frequencies where both are non-zero
- In reality, this overlap is qualified further:
  - The contribution of one fluorochrome into the readings of the other depends on the contributed energy into the other fluorochrome's detection filter
  - Thus, if the filters can be selected at frequencies where there is no overlap, there would be no correlation in the measurements
    - Problematic for several reasons
  - Otherwise, the contributions need to be undone numerically
    - Either on the instrument or computationally after-the-fact



Source: http://crl.berkeley.edu/compensation.html (From Operation Principles, Beckman Coulter Corp.)

Spectral Overlap

- In essence, recovering the intensities due to the individual fluorochromes amounts to a simple linear algebra problem
- Let
  - I_{FITC} denote the intensity of FITC
  - I_{PE} denote the intensity of PE
- Suppose that
  - a fraction A of FITC intensity leaks into the PE detector FL2
  - and a fraction B of PE intensity leaks into the FITC detector FL1
- This corresponds to a linear system of two equations

$$\text{FL1} = I_{FITC} + B\, I_{PE}$$
$$\text{FL2} = A\, I_{FITC} + I_{PE}$$

to be solved for two unknowns, I_{FITC} and I_{PE}

- The solution is then identified as

$$I_{FITC} = (1 - A\, B)^{-1} (\text{FL1} - B\, \text{FL2})$$
$$I_{PE} = (1 - A\, B)^{-1} (\text{FL2} - A\, \text{FL1})$$

- Note:
  - This requires the knowledge of the fractions A and B
  - But this depend on
    - the specifics of antibody conjugation of the fluorochromes and
    - the properties/settings of the optical detection system
    - as well as their respective emission spectra

Compensation

- In the absence of an exact knowledge of *A* and *B*, an approximate solution for compensation has been derived
  - primarily for two-color studies
- Observation:

  When *A* and *B* are relatively smaller than 1, $A\,B << 1$

  $$\rightarrow 1/(1 - A\,B) \approx 1$$
- As a result,

$$I_{FITC} \approx (FL1 - B\,FL2)$$
$$I_{PE} \approx (FL2 - A\,FL1)$$

where *A* and *B* now can be determined by trial and error to meet a specific criterion for successful compensation

Two-color Compensation

- Protocol for two-color compensation for FITC – PE:
  - Run --, -+ and +- cells from the flow cytometer
    - ++ samples can be present, but are not useful for compensation purposes
  - Adjust *A* so that (FL2 – *A* FL1) for -- cells is similar to (FL2 – *A* FL1) for +- cells
  - Adjust *B* so that (FL1 – *B* FL2) for -- cells is similar to (FL1 – *B* FL2) for -+ cells
- Notes:
  - Compensation starts with *A* because the spillover of PE into FL1 is generally much smaller than the spillover of FITC into FL2
    - Thus, FL1 is closer to the unknown I_{FITC} than FL2 is to the unknown I_{PE}
    - This allows using $I_{FITC} \approx$ FL1 when computing the correct I_{PE}
  - Calculation of *B* can then use the estimated using

    (FL1 – *B* (FL2 – *A* FL1))

    to determine the corrected I_{PE}

Two-color Compensation

- Notes (continued):
  - The most critical component of the compensation protocol is to determine when the fluorescence intensities are properly compensated
    - When the negatives in the +- cells match the corresponding negatives in the -- cells in distribution
  - Three options are available
    - Match the maximum fluorescence intensities
      - Unreliable as it represents an extreme value
    - Match the mean fluorescence intensities
      - Gets distorted by the logarithmic data transformation
    - Match the median fluorescence intensities
      - Generally better; statistically stable and unaffected by the logarithm

Two-color Compensation

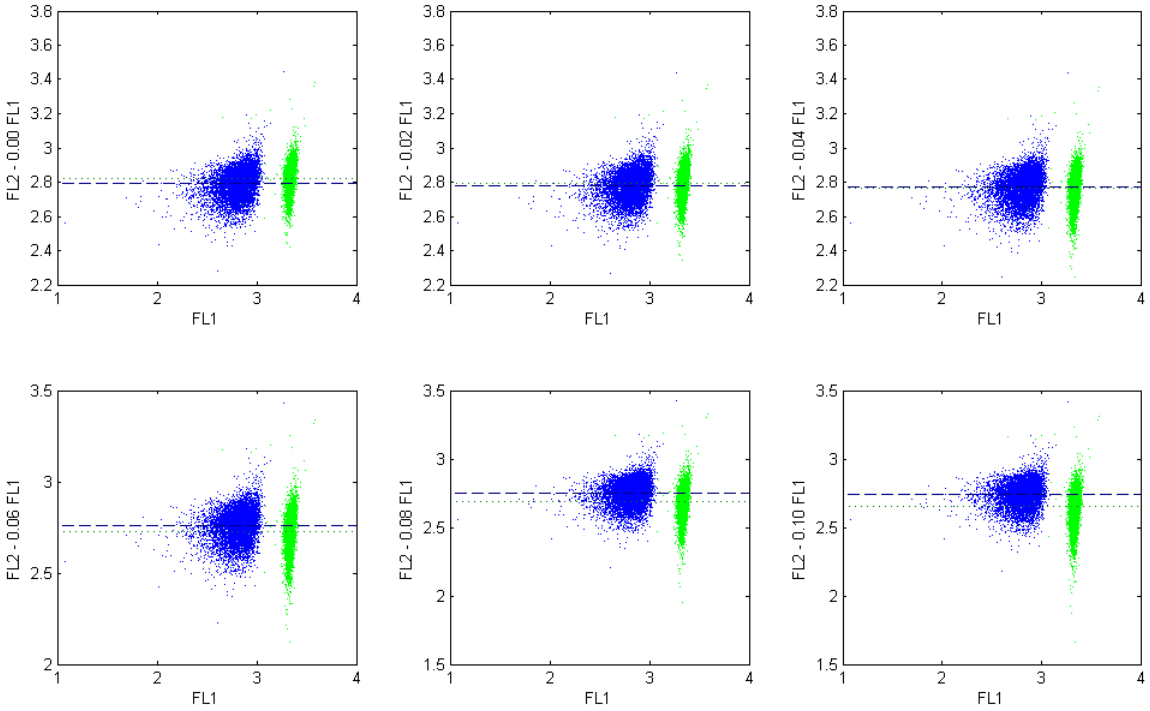
- Example:
  - Dataset provided by the reference

    I.Sugár, J. González-Lergier, and Stuart C. Sealfon, “Improved Compensation in Flow Cytometry by Multi-Variable Optimization,” Cytometry, Part A, Volume 79A, Issue 5, pages 356–360, 2011
    - 5-stained dendritic cells with 5 single-stained and 8 multi-stained controls
      - FITC, PE, Pacific Blue, PE-Cy7-A, and APC-A
  - The compensation protocol used to the FITC-PE compensation

Two-color Compensation

- Example (continued):
  - Adjusting *A* using -- and +- cells

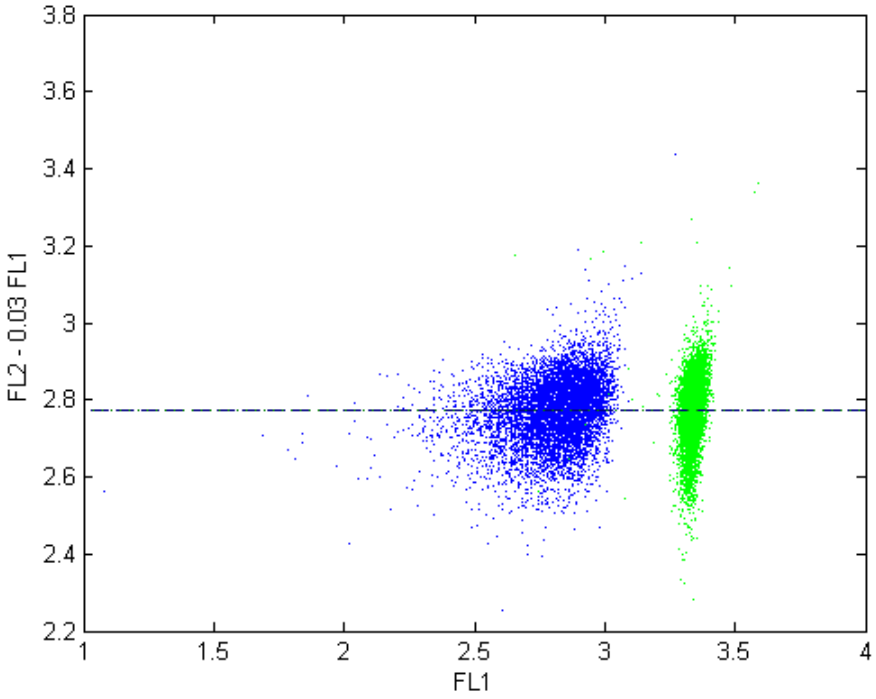
under-compensated



over-compensated

Two-color Compensation

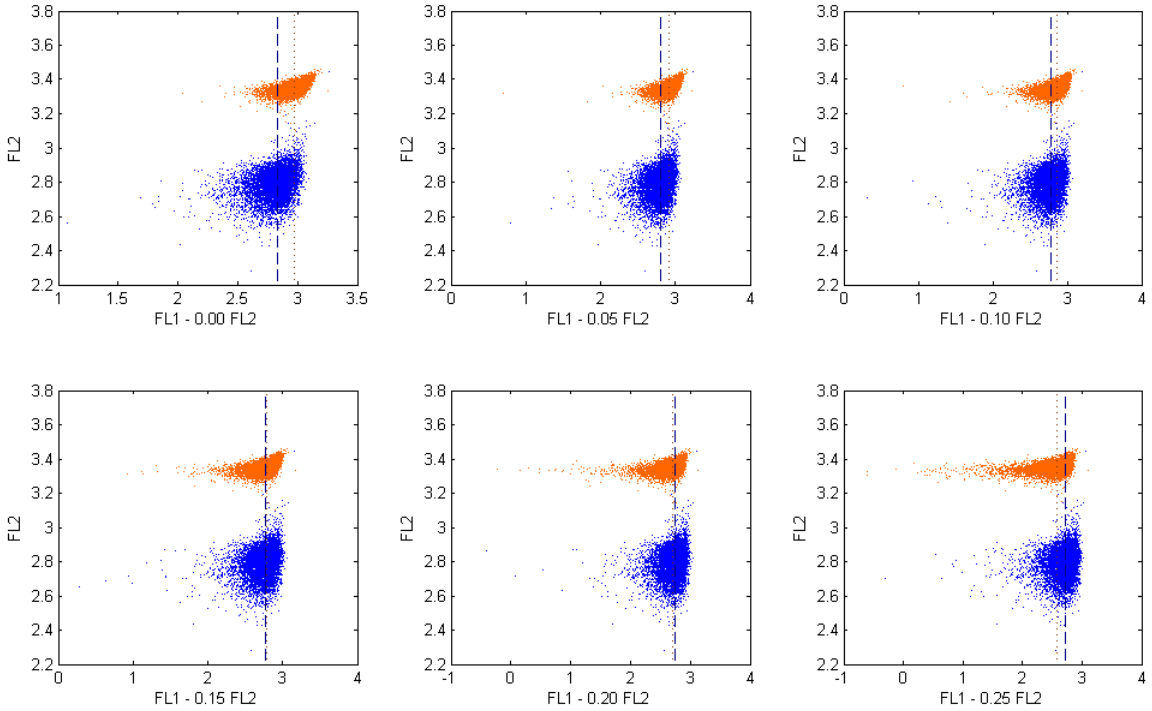
- Example (continued):
  - Adjusting *A* (final)



Two-color Compensation

- Example (continued):
  - Adjusting *B* using -- and -+ cells

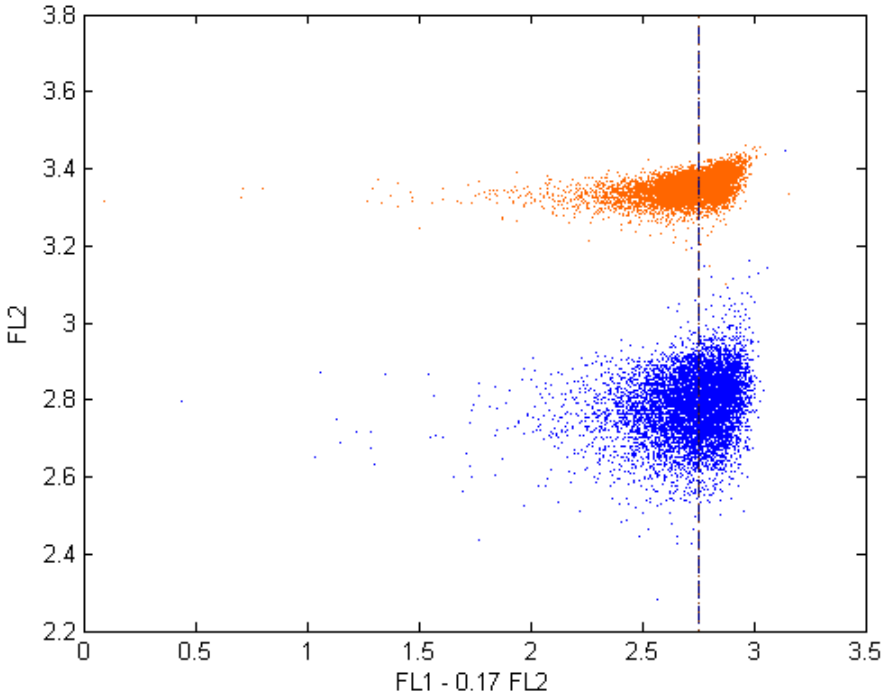
under-compensated



over-compensated

Two-color Compensation

- Example (continued):
  - Adjusting *B* (final)



Two-color Compensation

- Remarks:
  - This example represents an idealized compensation instance
    - Separate --, -+, and +- cell populations run through the flow cytometer separately
  - This avoids the issue of determining the positively and negatively stained cell populations on a single dataset
    - These cell populations may not exist or be easily differentiated
  - Additional issues need to be taken into account regarding the flow cytometer settings
    - Detector voltage gains, etc.
  - Subtraction of the fluorescence intensities can produce negative intensity values for the compensated channels
    - Aberration of the measurements

Compensation for Higher Numbers of Colors

- Three-color compensation:
  - A protocol exists to extend the two-color compensation routine to three colors
  - It assumes the fluorochromes for FL1 and FL3 do not leak into each other's detectors
  - Thus, after a two-color compensation on FL1 and FL2, it invokes a second two-color compensation for FL2 and FL3

Compensation for Higher Numbers of Colors

- Three-color compensation protocol:
  Source: http://flowcyt.salk.edu/howto/compensation/compensation-howto.html

- Make a mixture of approximately equal parts of unstained cells, green-only cells and orange-only cells (U/G/O mix).
- Make a mixture of approximately equal parts of unstained cells, orange-only cells and red-only cells (U/O/R mix).
- Run U/G/O mix on cytometer.
- Adjust FSC/SSC to bring cells on scale and set a gate around the population in which you are interested (possibly all cells).
- Construct a dot plot showing FL2 (y) versus FL1 (x). Set this plot to display only events in the above gate.
- Construct a dot plot showing FL3 (y) versus FL2 (x). Set this plot to display only events in the above gate.
- Set FL1, FL2 and FL3 amplifiers to logarithmic mode.
- Adjust the voltage on the FL1 detector to position the negative cell population at about FL1=100, but keep the bright FL1 population on scale.
- Adjust the voltage on the FL2 detector to position the negative cell population at about FL2=100, but keep the bright FL2 population on scale.
- Run U/O/R mix on cytometer. Refer to FL3 versus FL2 plot.
- Adjust the voltage on the FL3 detector to position the negative cell population at about FL3=100, but keep the bright FL3 population on scale.
- Adjust the FL3-%FL2 compensation control to bring the orange-only cell population downwards until the FL3 median of this population is approximately the same as the FL3 median of the negative cells.
- Adjust the FL2-%FL3 compensation control to bring the red-only cell population leftwards until the FL2 median of this population is approximately the same as the FL2 median of the negative cells.
- Run U/G/O mix on cytometer. Refer to FL2 versus FL1 plot.
- Adjust the FL2-%FL1 compensation control to bring the green-only cell population downwards until the FL2 median of this population is approximately the same as the FL2 median of the negative cells.
- Adjust the FL1-%FL2 compensation control to bring the orange-only cell population leftwards until the FL1 median of this population is approximately the same as the FL1 median of the negative cells.
- Some cytometers (e.g. LSR) have an extra pair of compensation controls for FL1 and FL3. You may wish to create a dotplot of FL3 versus FL1 and adjust FL3-%FL1 and FL1-%FL3 at this point.
- For most purposes it is adequate to set compensation "by eye" as in the 2-color example. If you wish to be as accurate as possible however, you must now draw regions around the 3 cell populations, display statistics and make fine adjustments to the compensation controls in order to equalize the median fluorescence values.
- Compensation is now set correctly.
- **If you subsequently change the voltage on any of the three detectors, or change the detection filters, you must repeat the compensation procedure!**

Compensation for Higher Numbers of Colors

- Compensation for more than three colors:
  - High color compensation requires multiply labeled control cells
    - The binding of the antibodies to the cells in the suspension must be known
  - The manual approach is inadequate due to the complex effect of compensating one fluorochrome onto the others
  - Especially when tandem dyes are used, each experiment must be followed by a separate compensation routine carried out from scratch

Compensation for Higher Numbers of Colors

- Method by linear algebra:
  - Given control samples of
    - unstained cells and
    - singly-stained cells
  - Compute the spill-over coefficients $C_{j,k}$ via

$$C_{j,k} = (\mu_{j,k} - \mu_{j,0})/(\mu_{k,k} - \mu_{k,0})$$

    - $\mu_{j,k}$: average intensity in FLj in samples stained by the k'th fluorochrome
    - $\mu_{j,0}$: average intensity in FLj in unstained samples
    - $\mu_{k,k}$: average intensity in FLk in samples stained by the k'th fluorochrome
    - $\mu_{k,0}$: average intensity in FLk in unstained samples

    to measure the contribution of the k'th fluorochrome into the detected intensity for the j'th fluorochrome
  - The compensated intensities to be obtained by matrix inversion

$$[I_1\ I_2\ \ldots\ I_K]^T = C^{-1}\ [\mathrm{FL1\ FL2}\ \ldots\ \mathrm{FL}K]^T$$

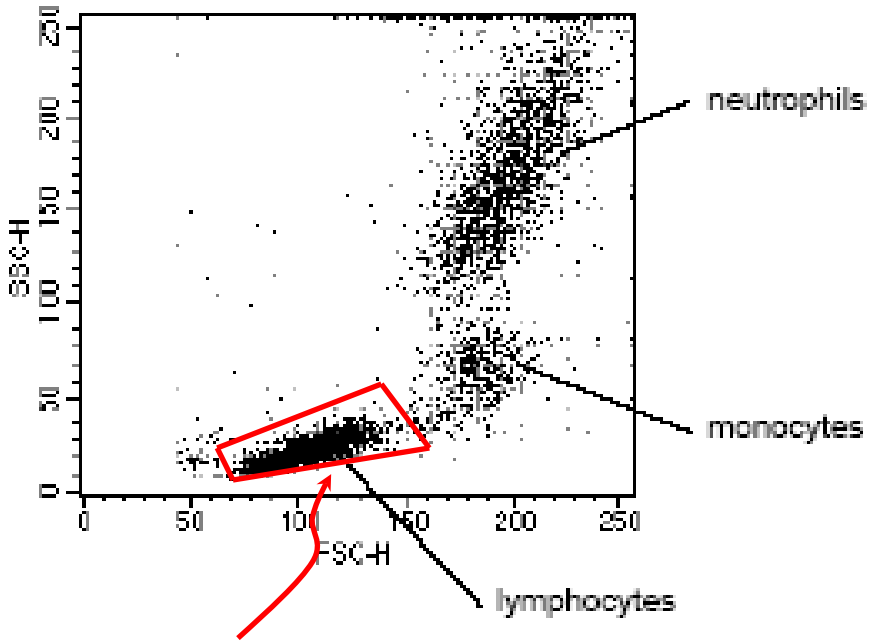
- Note: This method disregards the interaction between multiple fluorochromes that jointly contribute to the measured intensities

Gating

- Once the intensities are properly compensated, different cell subtypes occupy different regions in the observation space
- The staining properties of a specific cell subtype can be studied separately by identifying them on the scatter plots
  - Desired cells are identified as those that reside in a specific region on the scatter plot
    - with respect to two parameters that allow for their detection
  - Once the cells satisfying the selection criterion are identified, their staining properties with respect to the other biomarkers can be studied separately
    - without interference from the other cells

Gating

- Example: Selection of lymphocytes
  - A scatter plot of leukocytes using forward scatter versus side scatter allows selective identification of the lymphocytes from the neutrophils and the monocytes
  - A region drawn around the lymphocytes eliminates the non-lymphocytes from the subsequent analysis



lymphocyte gate

Source: http://www.med.umich.edu/flowcytometry/training/lessons/lesson1/

Summary

- Correct compensation is critical for an accurate analysis of the flow cytometry data
  - Inaccurate compensation can prevent the identification of cell subtypes of interest
- Further improvement can be achieved on the method based on matrix algebra
  - by taking into account the joint contribution of multiple fluorochromes to the detected intensities
    - Optimization-based techniques proposed in the literature
- Following adequate compensation, manual gating allows studying the staining patterns of cell subtypes
  - Large body of literature exists on automated gating
  - But it requires standardization of the multicolor flow data beyond proper compensation

  ➔ Flow data normalization