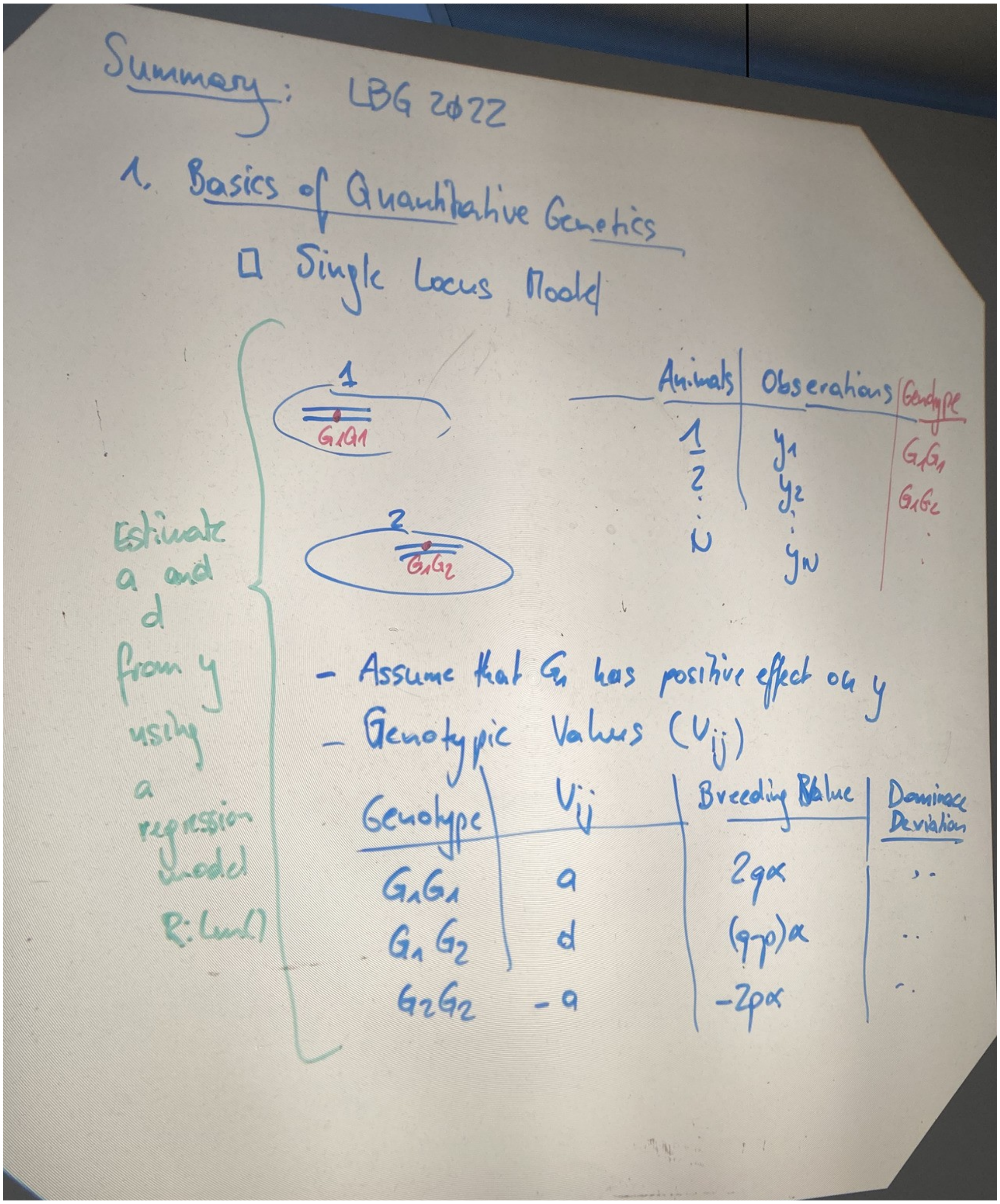
## OHP Picture 1



Summary: LBG 2022

1. Basics of Quantitative Genetics

☐ Single Locus Model

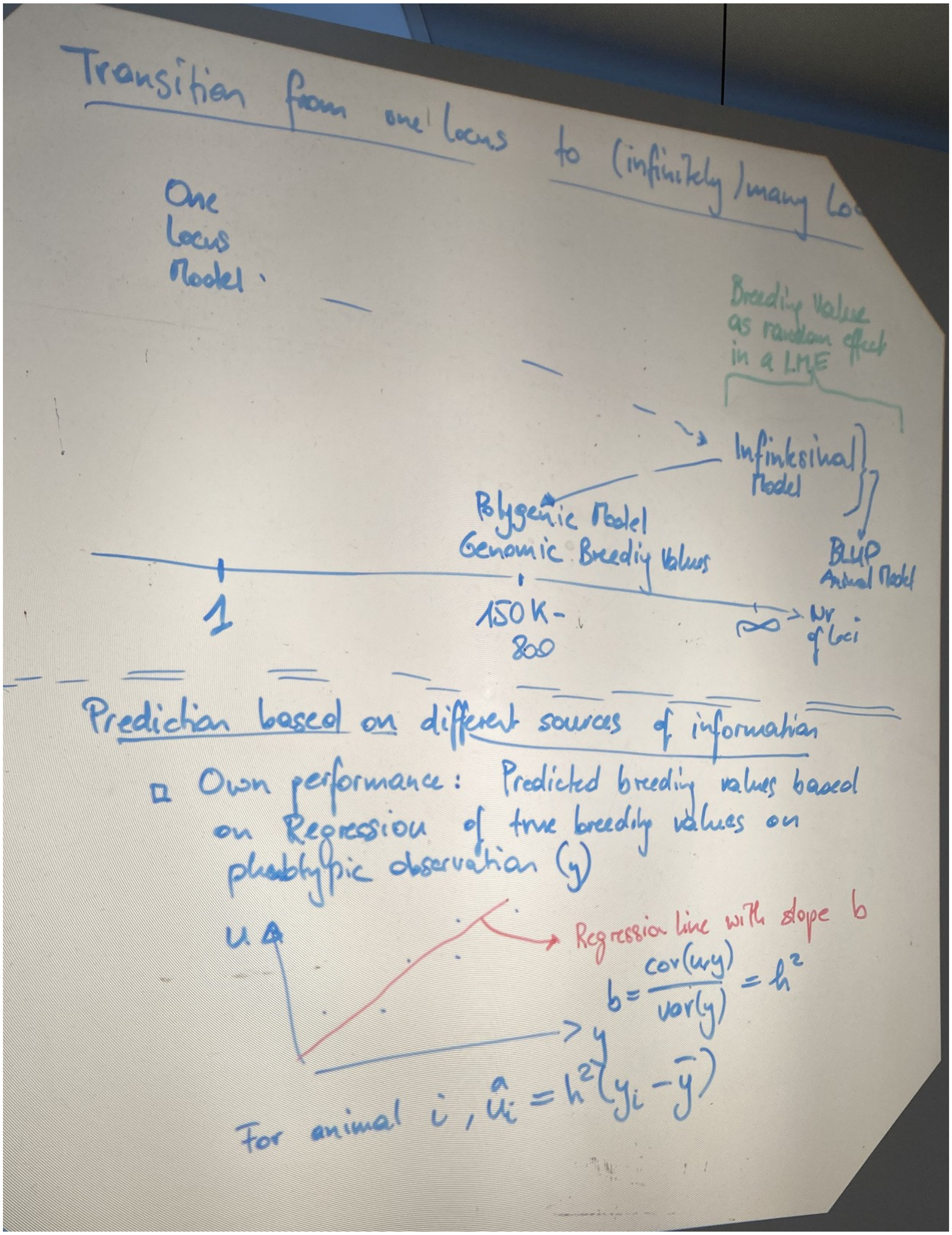
| Animals | Observations | Genotype |
|---|---|---|
| 1 | $y_1$ | $G_1G_1$ |
| 2 | $y_2$ | $G_1G_2$ |
| ⋮ | ⋮ | ⋮ |
| N | $y_N$ | |

Estimate a and d from y using a regression model R: lm()

- Assume that $G_1$ has positive effect on y
- Genotypic Values ($V_{ij}$)

| Genotype | $V_{ij}$ | Breeding Value | Dominance Deviation |
|---|---|---|---|
| $G_1G_1$ | a | $2q\alpha$ | .. |
| $G_1G_2$ | d | $(q-p)\alpha$ | .. |
| $G_2G_2$ | -a | $-2p\alpha$ | .. |

## OHP Picture 2



Transition from one locus to (infinitely) many Lo

One Locus Model

Breeding Value as random effect in a LME

Infinitesimal Model

Polygenic Model Genomic Breeding Values

BLUP Animal Model

1

150K - 800

$\infty$

Nr of loci

Prediction based on different sources of information

- ☐ Own performance: Predicted breeding values based on Regression of true breeding values on phaenotypic observation (y)

u

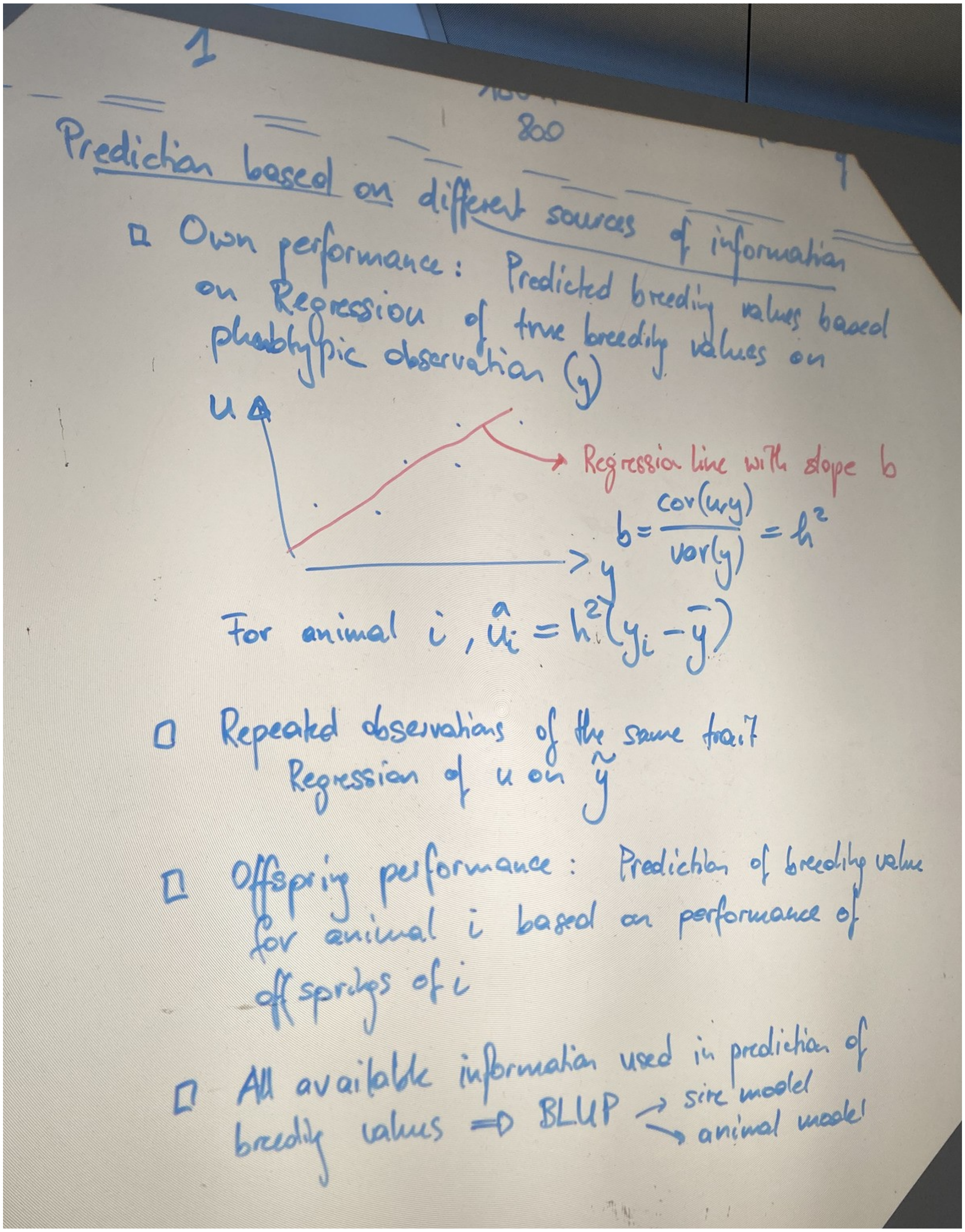
y

Regression line with slope b

$$b = \frac{cov(u,y)}{var(y)} = h^2$$

For animal i, $\hat{u}_i = h^2(y_i - \bar{y})$

## OHP Picture 3



1

800

Prediction based on different sources of information

☐ Own performance: Predicted breeding values based on Regression of true breeding values on phenotypic observation (y)

$$b = \frac{cov(u,y)}{var(y)} = h^2$$

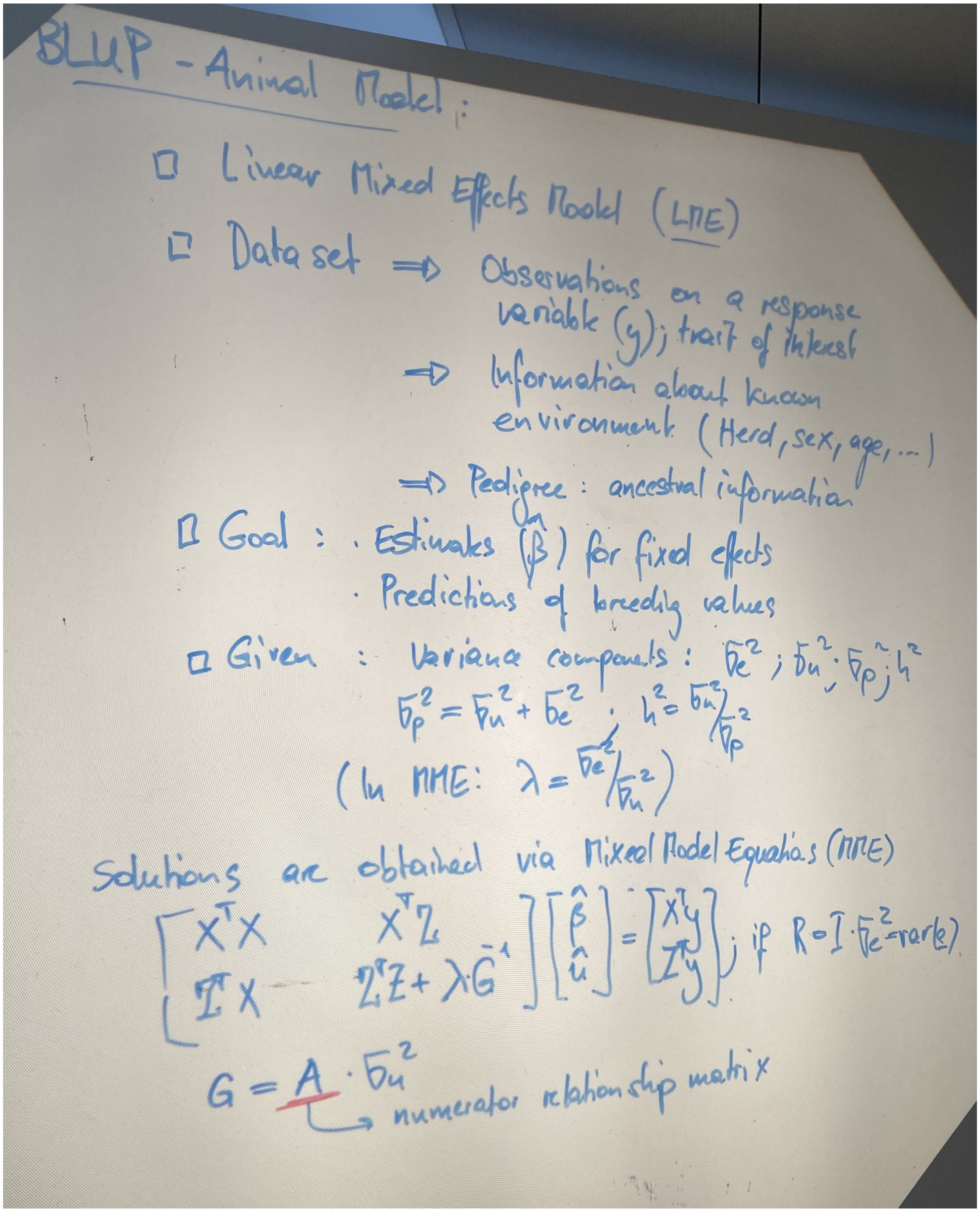
For animal i, $\hat{u}_i = h^2(y_i - \bar{y})$

☐ Repeated observations of the same trait
Regression of u on $\tilde{y}$

☐ Offspring performance: Prediction of breeding value for animal i based on performance of offsprings of i

☐ All available information used in prediction of breeding values $\Rightarrow$ BLUP → sire model, → animal model

## OHP Picture 4



# BLUP - Animal Model:

- ☐ Linear Mixed Effects Model (LME)
- ☐ Data set ⇒ Observations on a response variable (y); trait of interest
  - ⇒ Information about known environment (Herd, sex, age, ...)
  - ⇒ Pedigree: ancestral information
- ☐ Goal:
  - Estimates ($\beta$) for fixed effects
  - Predictions ($u$) of breeding values
- ☐ Given: Variance components: $\sigma_e^2$; $\sigma_u^2$; $\sigma_p^2$; $h^2$

$$\sigma_p^2 = \sigma_u^2 + \sigma_e^2 \; ; \; h^2 = \sigma_u^2 / \sigma_p^2$$

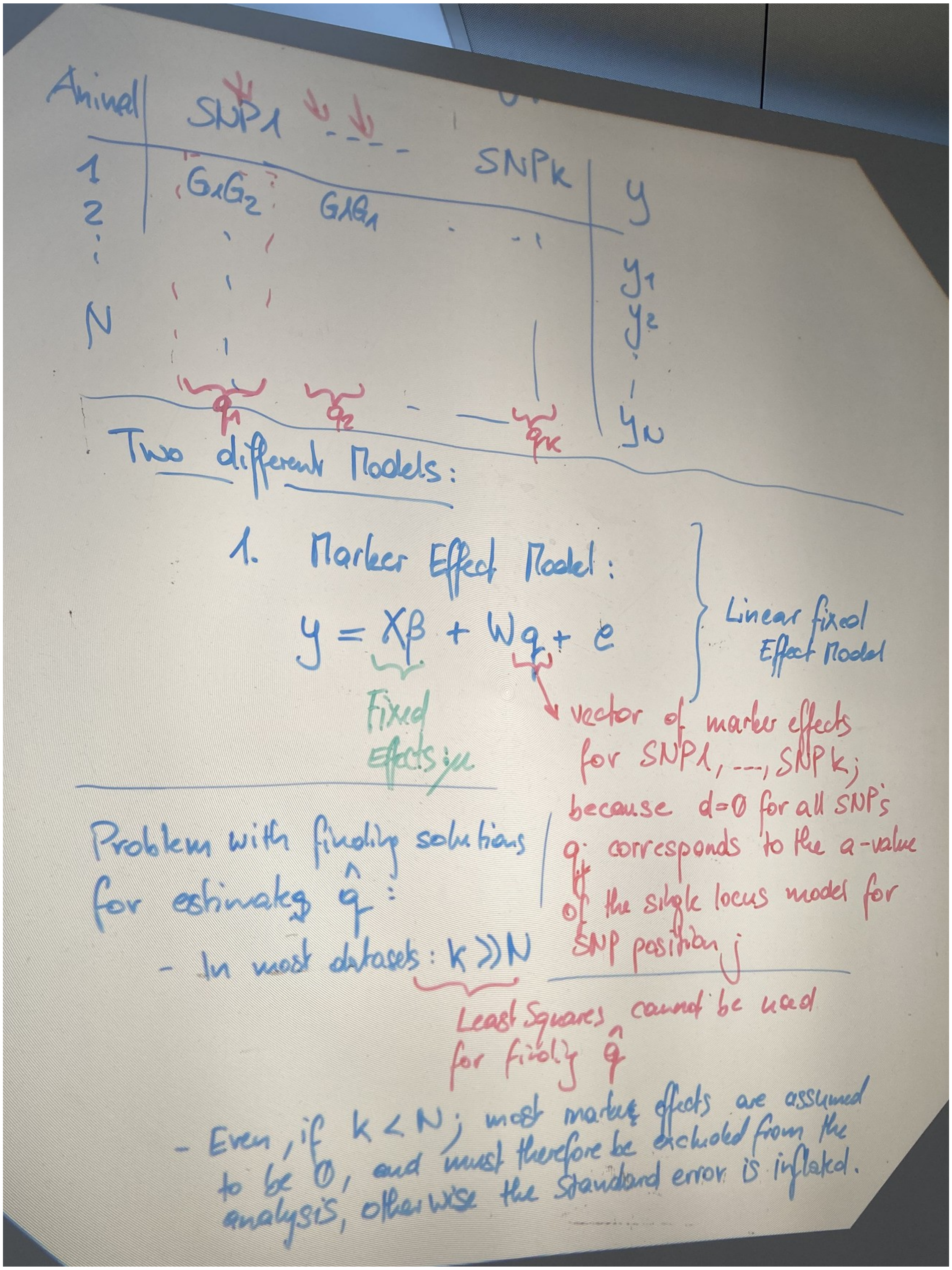
$$\left(\text{In MME: } \lambda = \sigma_e^2 / \sigma_u^2\right)$$

Solutions are obtained via Mixed Model Equations (MME)

$$\begin{bmatrix} X^T X & X^T Z \\ Z^T X & Z^T Z + \lambda \cdot G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X^T y \\ Z^T y \end{bmatrix} ; \text{ if } R = I \cdot \sigma_e^2 = var(e)$$

$$G = \underbrace{A}_{\text{numerator relationship matrix}} \cdot \sigma_u^2$$

**OHP Picture 5**



| Animal | SNP1 | - - - - | SNPk | y |
|---|---|---|---|---|
| 1 | $G_1G_2$ | $G_1G_1$ | | $y_1$ |
| 2 | | | | $y_2$ |
| ⋮ | | | | ⋮ |
| N | | | | $y_N$ |
| | $q_1$ | $q_2$ - - | $q_k$ | |

Two different Models:

1. Marker Effect Model:

$$y = \underbrace{X\beta}_{\text{Fixed Effects: } \mu} + \underbrace{Wq}_{} + e$$

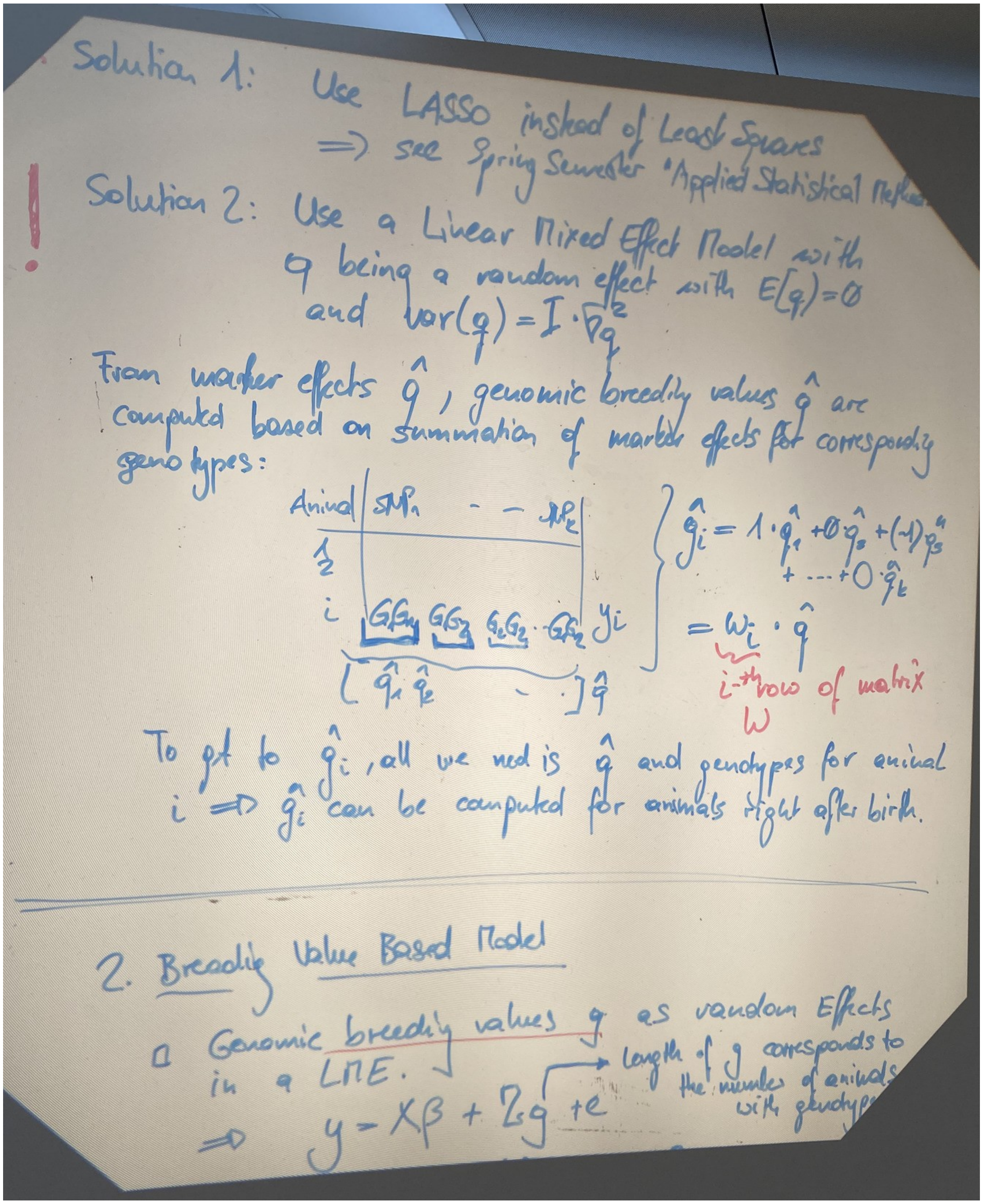
Linear fixed Effect Model

vector of marker effects for SNP1, ---, SNPk; because $d=0$ for all SNP's $q_j$ corresponds to the a-value of the single locus model for SNP position $j$

Problem with finding solutions for estimates $\hat{q}$:

- In most datasets: $\underbrace{k \gg N}_{}$ Least Squares cannot be used for finding $\hat{q}$
- Even, if $k < N$; most marker effects are assumed to be 0, and must therefore be excluded from the analysis, otherwise the standard error is inflated.

## OHP Picture 6



Solution 1: Use LASSO instead of Least Squares

$\Longrightarrow$ see Spring Semester "Applied Statistical Meth

!

Solution 2: Use a Linear Mixed Effect Model with $q$ being a random effect with $E[q] = 0$ and $var(q) = I \cdot \sigma_q^2$

From marker effects $\hat{q}$, genomic breeding values $\hat{g}$ are computed based on summation of marker effects for corresponding genotypes:

$$\begin{array}{c|cccc|c} \text{Animal} & SNP_1 & & \cdots & SNP_k & \\ \hline 1 & & & & & \\ 2 & & & & & \\ i & \underbrace{G_1G_1} & \underbrace{G_1G_2} & \underbrace{G_2G_2} \cdots & G_1G_2 & y_i \\ & [\ \hat{q}_1 & \hat{q}_2 & & \cdots\ ] & \hat{q} \end{array}$$

$$\left.\right\} \hat{g}_i = 1 \cdot \hat{q}_1 + 0 \cdot \hat{q}_2 + (-1) \hat{q}_3 + \cdots + 0 \cdot \hat{q}_k$$

$$= \underline{w_i} \cdot \hat{q}$$

$i$-th row of matrix $W$

To get to $\hat{g}_i$, all we need is $\hat{q}$ and genotypes for animal $i$ $\Longrightarrow$ $\hat{g}_i$ can be computed for animals right after birth.

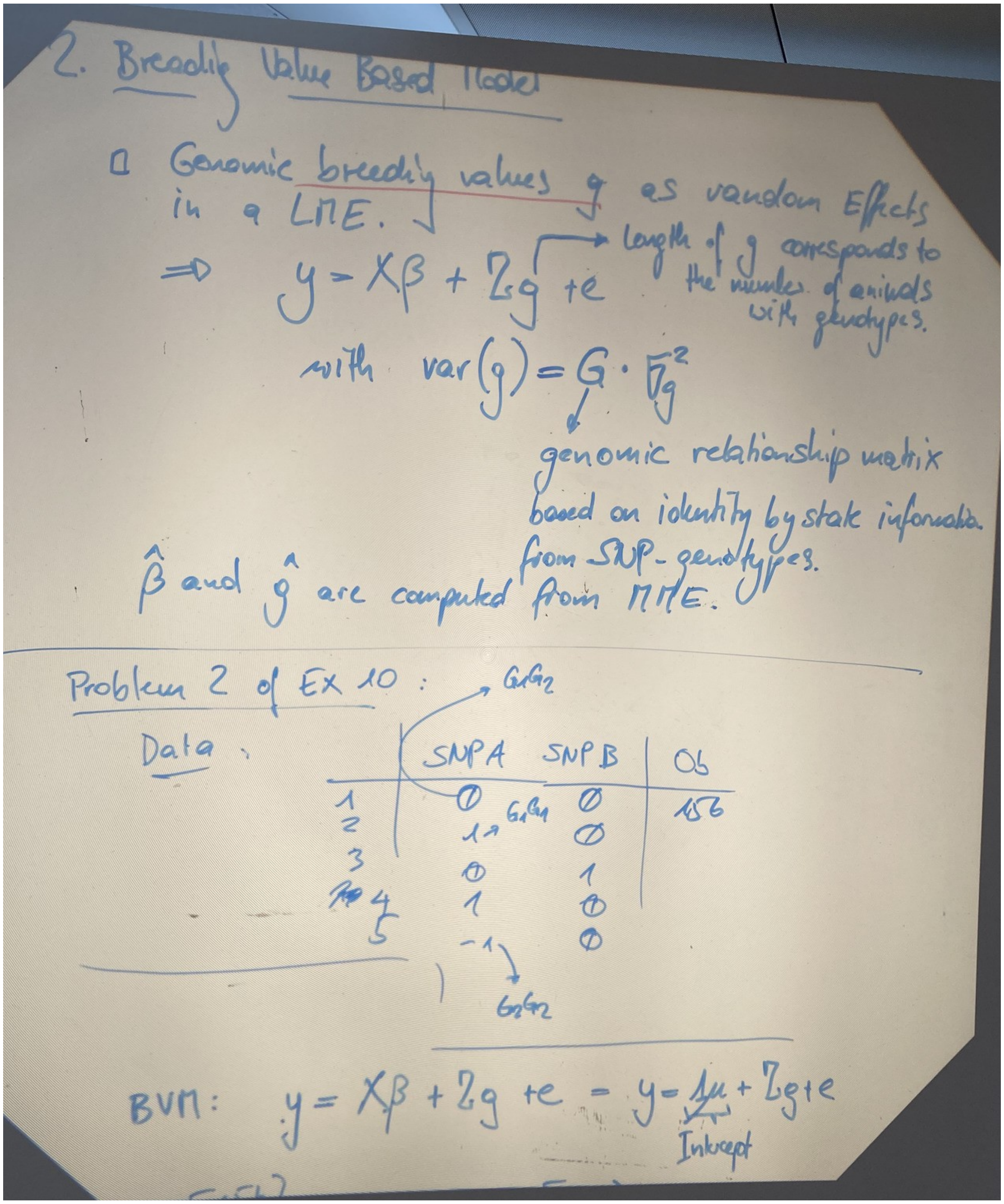
---

2. Breeding Value Based Model

- Genomic breeding values $g$ as random Effects in a LME.

$\Longrightarrow$ $y = X\beta + Zg + e$ $\rightarrow$ Length of $g$ corresponds to the number of animals with genotyp

**OHP Picture 7**



## 2. Breeding Value Based Model

- Genomic breeding values $g$ as random Effects in a LME.

$$\Rightarrow \quad y = X\beta + Zg + e$$

Length of $g$ corresponds to the number of animals with genotypes.

with $var(g) = G \cdot \sigma_g^2$

$G$: genomic relationship matrix based on identity by state information from SNP-genotypes.

$\hat{\beta}$ and $\hat{g}$ are computed from MME.

---

Problem 2 of Ex 10:

Data:

| | SNP A | SNP B | Ob |
|---|---|---|---|
| 1 | 0 | 0 | 156 |
| 2 | 1 | 0 | |
| 3 | 0 | 1 | |
| 4 | 1 | 0 | |
| 5 | -1 | 0 | |

$G_1G_2$

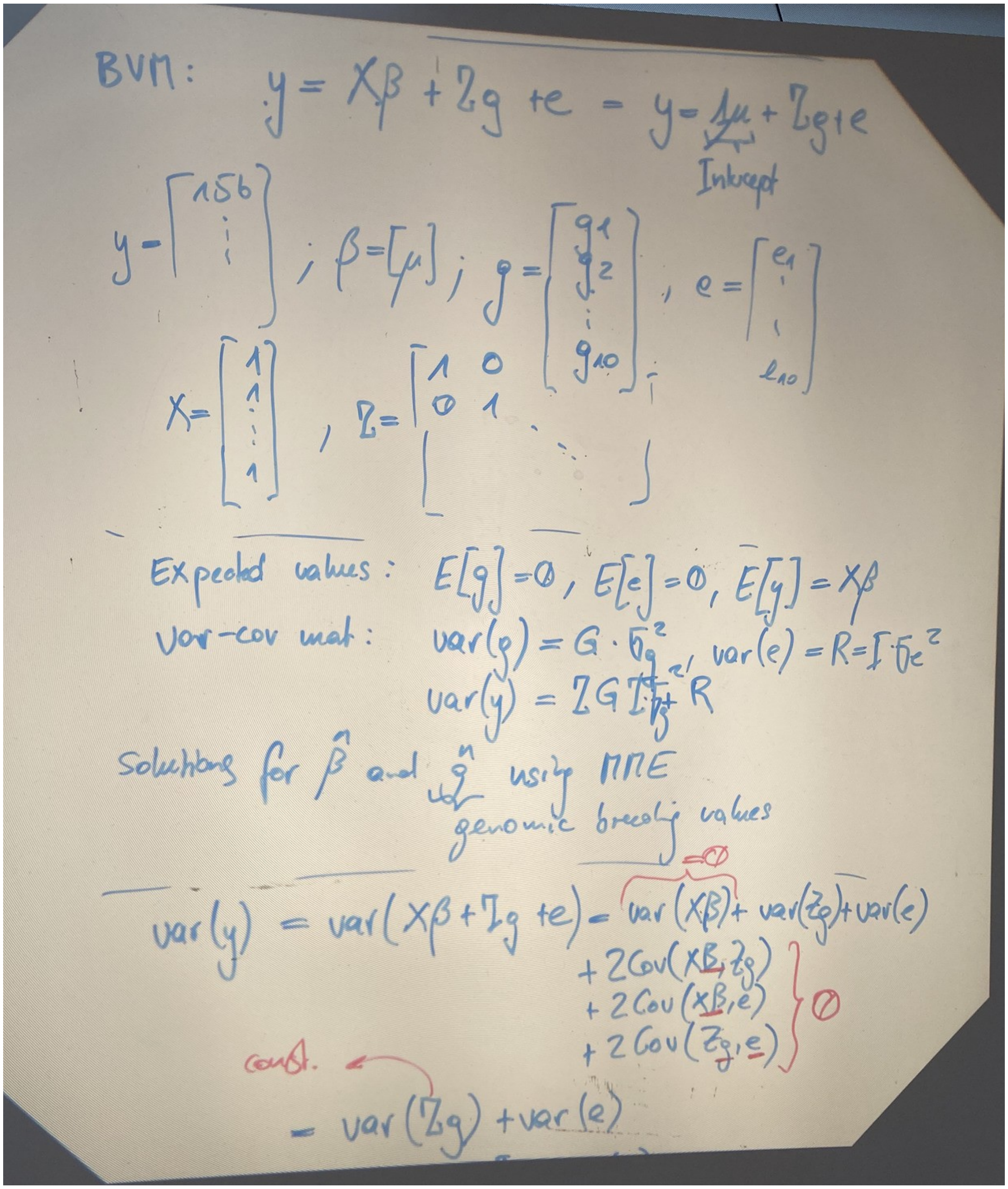
$G_1G_1$

$G_2G_2$

---

$$BVM: \quad y = X\beta + Zg + e = y = \underbrace{1\mu}_{\text{Intercept}} + Zg + e$$

## OHP Picture 8



BVM: $y = X\beta + Zg + e = y = \underbrace{1\mu}_{\text{Intercept}} + Zg + e$

$y = \begin{bmatrix} 156 \\ \vdots \\ \vdots \end{bmatrix}$; $\beta = [\mu]$; $g = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_{10} \end{bmatrix}$, $e = \begin{bmatrix} e_1 \\ \vdots \\ e_{10} \end{bmatrix}$;

$X = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$, $Z = \begin{bmatrix} 1 & 0 & & \\ 0 & 1 & & \\ & & \ddots & \\ & & & \end{bmatrix}$

Expected values: $E[g] = 0$, $E[e] = 0$, $E[y] = X\beta$

Var-cov mat: $var(g) = G \cdot \sigma_g^2$, $var(e) = R = I \cdot \sigma_e^2$

$var(y) = ZGZ'\sigma_g^2 + R$

Solutions for $\hat{\beta}$ and $\hat{g}$ using MME
or genomic breeding values

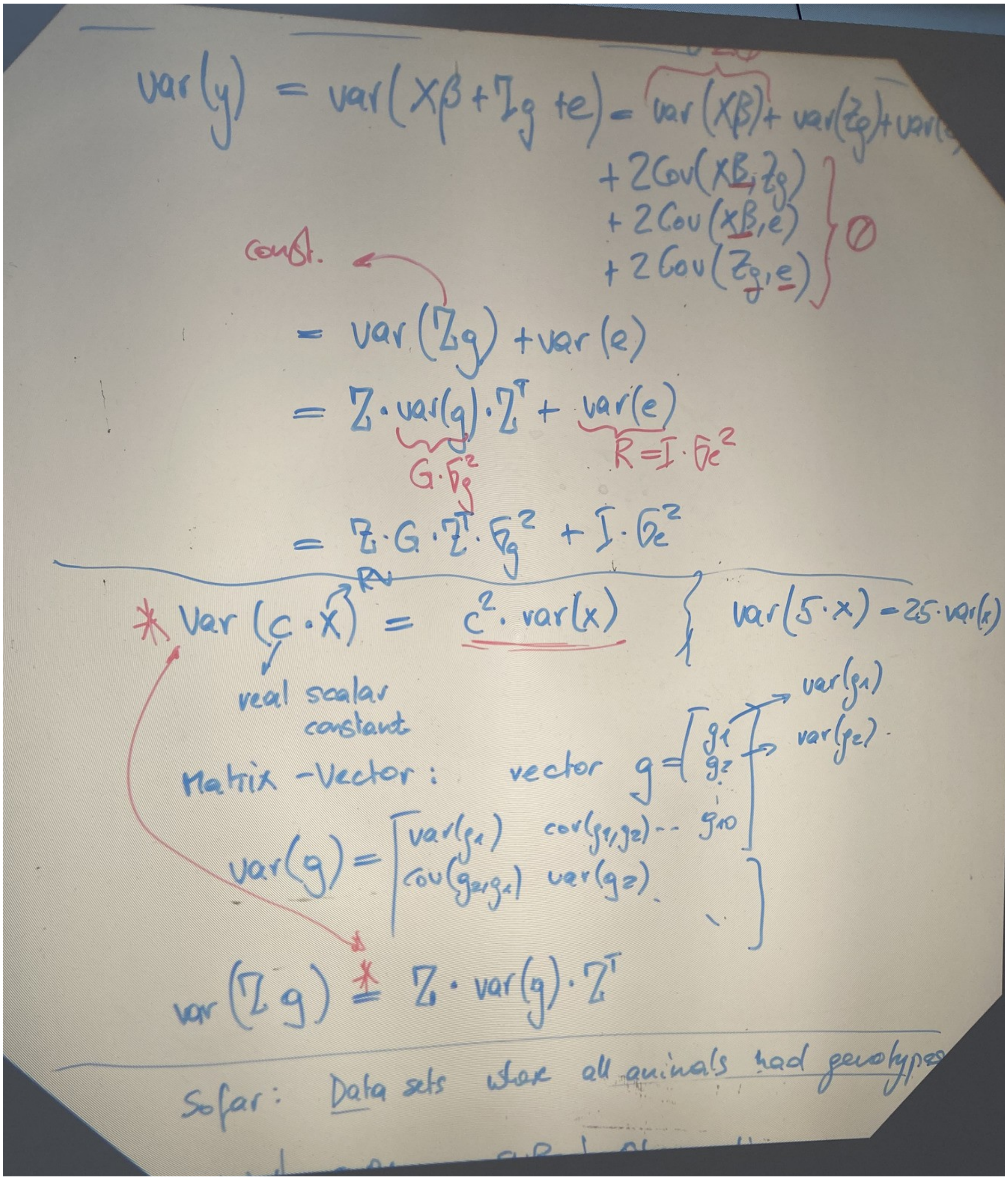
$var(y) = var(X\beta + Zg + e) = \overbrace{var(X\beta)}^{=0} + var(Zg) + var(e)$

$\left.\begin{array}{l} + 2Cov(X\underline{\beta}, Zg) \\ + 2Cov(X\underline{\beta}, e) \\ + 2Cov(Zg, \underline{e}) \end{array}\right\} 0$

const.

$= var(Zg) + var(e)$

## OHP Picture 9



$$\mathrm{var}(y) = \mathrm{var}(X\beta + Zg + e) = \mathrm{var}(X\beta) + \mathrm{var}(Zg) + \mathrm{var}(e)$$

$$\left.\begin{array}{l} + 2\,\mathrm{Cov}(X\beta, Zg) \\ + 2\,\mathrm{Cov}(X\beta, e) \\ + 2\,\mathrm{Cov}(Zg, e) \end{array}\right\} 0$$

const.

$$= \mathrm{var}(Zg) + \mathrm{var}(e)$$

$$= Z \cdot \underbrace{\mathrm{var}(g)}_{G \cdot \sigma_g^2} \cdot Z^T + \underbrace{\mathrm{var}(e)}_{R = I \cdot \sigma_e^2}$$

$$= Z \cdot G \cdot Z^T \cdot \sigma_g^2 + I \cdot \sigma_e^2$$

$$* \ \mathrm{Var}(c \cdot x) = c^2 \cdot \mathrm{var}(x) \quad \Big\} \quad \mathrm{var}(5 \cdot x) = 25 \cdot \mathrm{var}(x)$$

$c$: real scalar constant

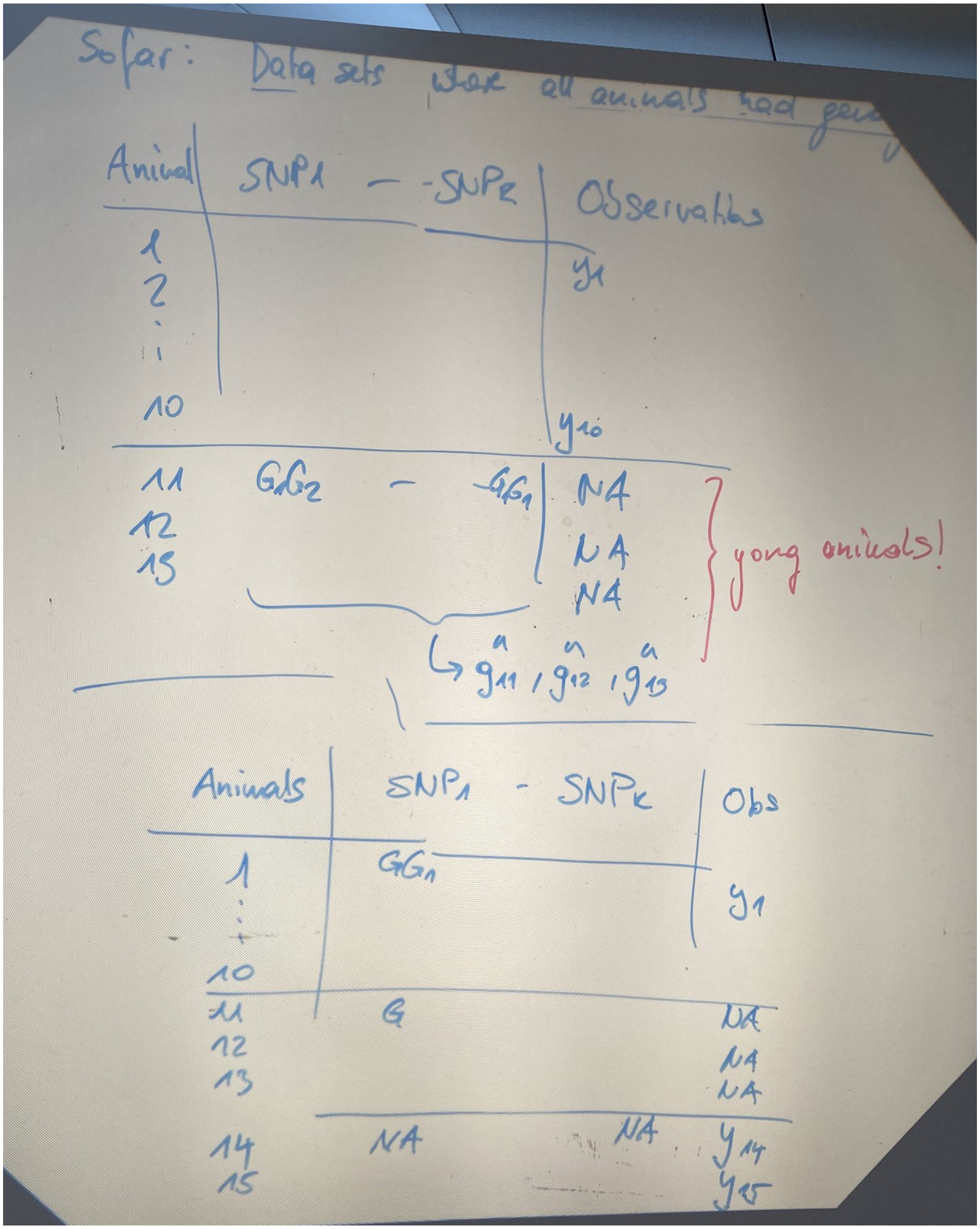
Matrix-Vector: vector $g = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_{10} \end{bmatrix}$ → $\mathrm{var}(g_1)$, $\mathrm{var}(g_2)$.

$$\mathrm{var}(g) = \begin{bmatrix} \mathrm{var}(g_1) & \mathrm{cov}(g_1, g_2) & \cdots \\ \mathrm{cov}(g_2, g_1) & \mathrm{var}(g_2) & \\ & & \ddots \end{bmatrix}$$

$$\mathrm{var}(Zg) \overset{*}{=} Z \cdot \mathrm{var}(g) \cdot Z^T$$

So far: Data sets where all animals had genotypes

## OHP Picture 10



Sofar: Data sets where all animals had gen

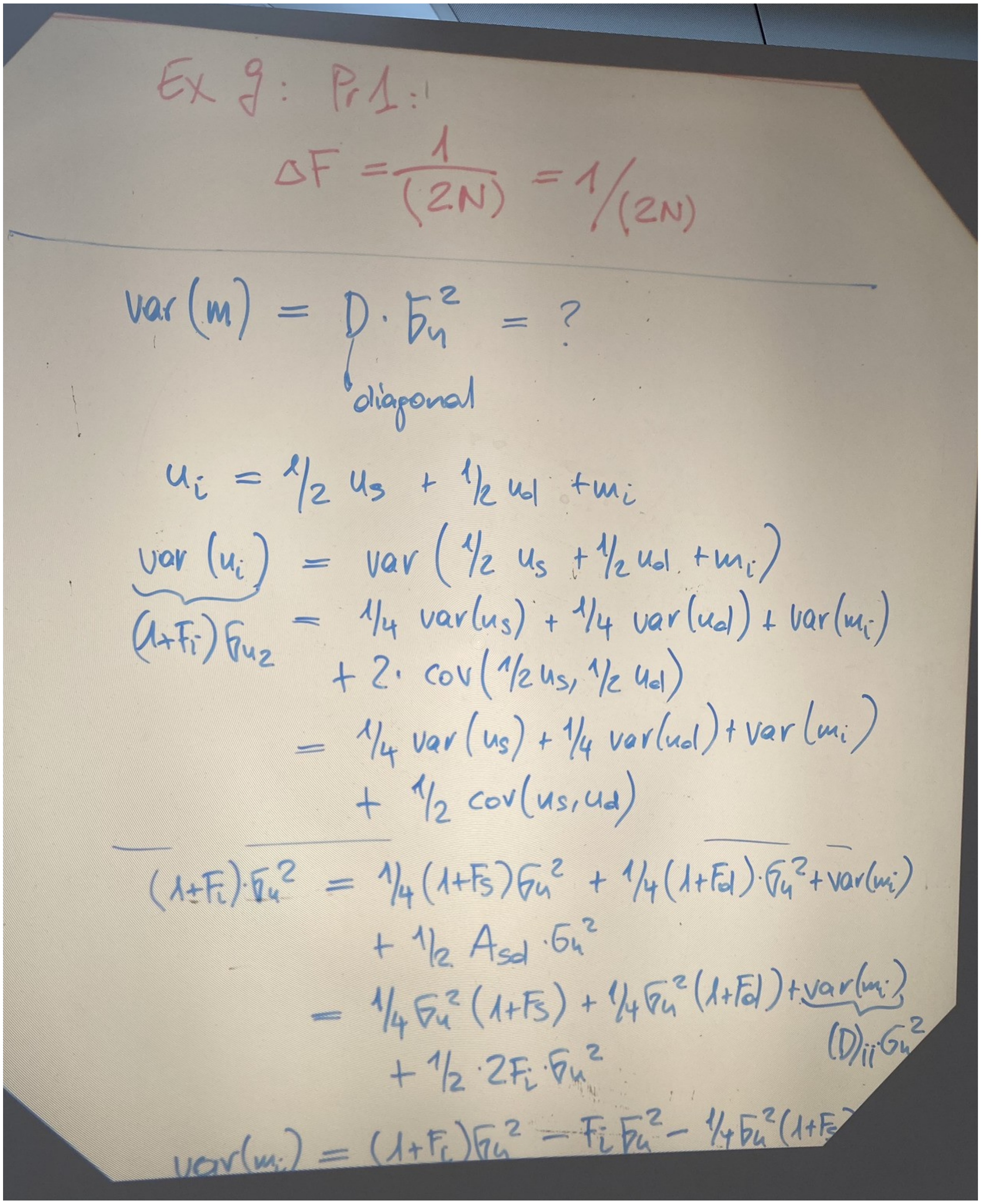
| Animal | SNP1 – – SNPk | Observations |
|---|---|---|
| 1 | | $y_1$ |
| 2 | | |
| ⋮ | | |
| 10 | | $y_{10}$ |
| 11 | $G_1G_2$ – $G_1G_1$ | NA |
| 12 | | NA |
| 13 | | NA |

$\underbrace{\qquad\qquad}$ ↳ $\hat{g}_{11}, \hat{g}_{12}, \hat{g}_{13}$

yong animals!

| Animals | $SNP_1$ - $SNP_k$ | Obs |
|---|---|---|
| 1 | $G_1G_1$ | $y_1$ |
| ⋮ | | |
| 10 | | |
| 11 | G | NA |
| 12 | | NA |
| 13 | | NA |
| 14 | NA NA | $y_{14}$ |
| 15 | | $y_{15}$ |

## OHP Picture 11



Ex 9: Pr 1:

$$\Delta F = \frac{1}{(2N)} = 1/(2N)$$

---

$$\mathrm{var}(m) = \underset{\text{diagonal}}{D} \cdot \sigma_u^2 = ?$$

$$u_i = 1/2\, u_s + 1/2\, u_d + m_i$$

$$\underbrace{\mathrm{var}(u_i)}_{(1+F_i)\sigma_{u2}} = \mathrm{var}(1/2\, u_s + 1/2\, u_d + m_i)$$

$$= 1/4\, \mathrm{var}(u_s) + 1/4\, \mathrm{var}(u_d) + \mathrm{var}(m_i)$$

$$+ 2 \cdot \mathrm{cov}(1/2\, u_s, 1/2\, u_d)$$

$$= 1/4\, \mathrm{var}(u_s) + 1/4\, \mathrm{var}(u_d) + \mathrm{var}(m_i)$$

$$+ 1/2\, \mathrm{cov}(u_s, u_d)$$

$$(1+F_i) \cdot \sigma_u^2 = 1/4 (1+F_s)\sigma_u^2 + 1/4 (1+F_d) \cdot \sigma_u^2 + \mathrm{var}(m_i)$$

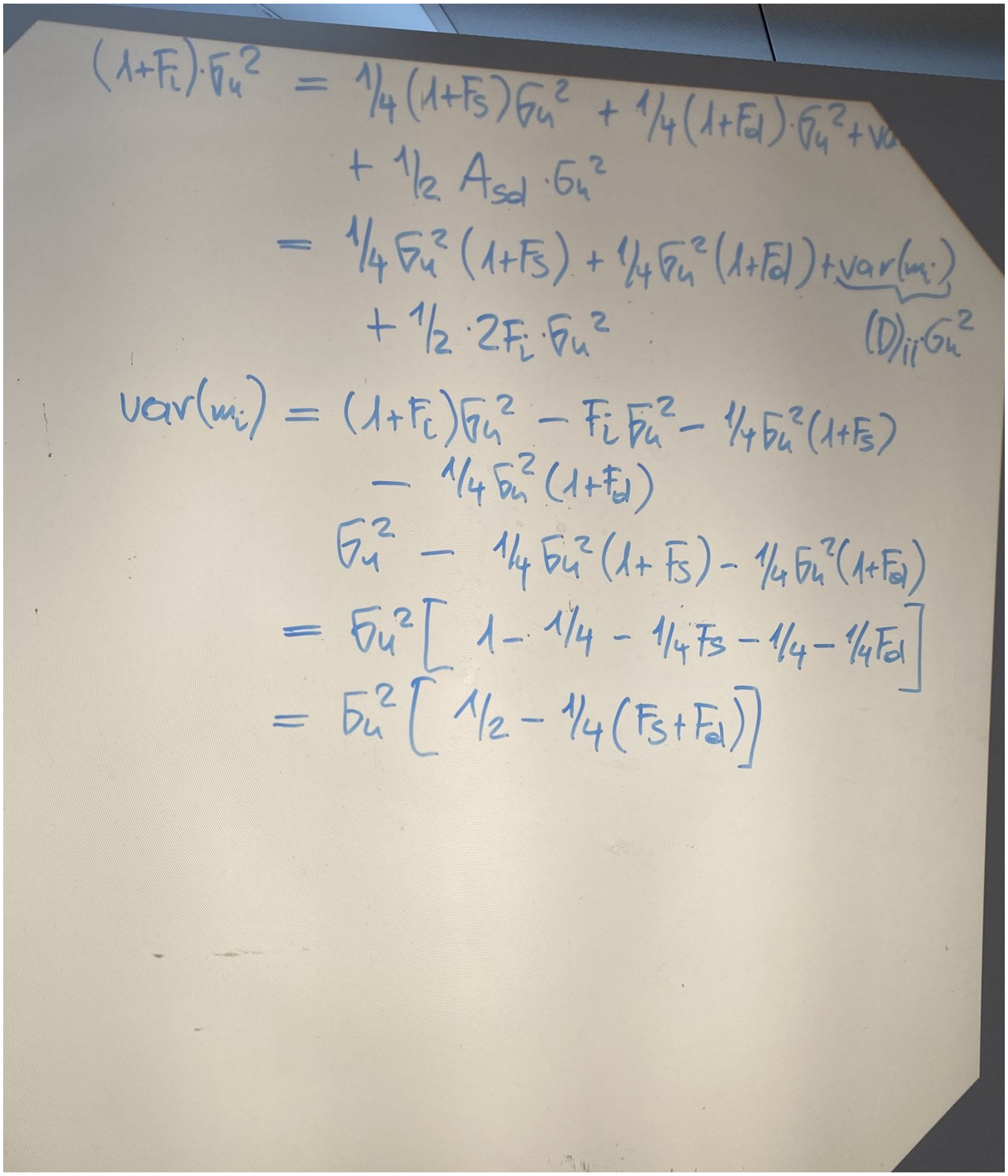
$$+ 1/2\, A_{sd} \cdot \sigma_u^2$$

$$= 1/4\, \sigma_u^2 (1+F_s) + 1/4\, \sigma_u^2 (1+F_d) + \underbrace{\mathrm{var}(m_i)}_{(D)_{ii} \cdot \sigma_u^2}$$

$$+ 1/2 \cdot 2F_i \cdot \sigma_u^2$$

$$\mathrm{var}(m_i) = (1+F_i)\sigma_u^2 - F_i \sigma_u^2 - 1/4\, \sigma_u^2 (1+F_s$$

## OHP Picture 12



$$(1+F_i)\cdot\sigma_u^2 = 1/4\,(1+F_s)\sigma_u^2 + 1/4\,(1+F_d)\cdot\sigma_u^2 + va$$

$$+ 1/2\, A_{sd}\cdot\sigma_u^2$$

$$= 1/4\,\sigma_u^2(1+F_s) + 1/4\,\sigma_u^2(1+F_d) + \underbrace{var(m_i)}_{(D)_{ii}\cdot\sigma_u^2}$$

$$+ 1/2\cdot 2F_i\cdot\sigma_u^2$$

$$var(m_i) = (1+F_i)\sigma_u^2 - F_i\sigma_u^2 - 1/4\,\sigma_u^2(1+F_s)$$

$$- 1/4\,\sigma_u^2(1+F_d)$$

$$\sigma_u^2 - 1/4\,\sigma_u^2(1+F_s) - 1/4\,\sigma_u^2(1+F_d)$$

$$= \sigma_u^2\left[1 - 1/4 - 1/4\,F_s - 1/4 - 1/4\,F_d\right]$$

$$= \sigma_u^2\left[1/2 - 1/4\,(F_s+F_d)\right]$$