Proof of the Natural Join Expression

Let $T_1$ and $T_2$ be tables whose schemas are $T_1$ and $T_2$, respectively. Let $A_1, ..., A_d$ be their common attributes. Prove:

$$T_1 \bowtie T_2 = \Pi_S(\sigma_P(T_1 \times T_2))$$

where $S = (S_1 - S_2) \cup \{T_1.A_1, ..., T_1.A_d\} \cup (S_2 - S_1)$ and $P = (T_1.A_1 = T_2.A_1 \land ... \land T_1.A_d = T_2.A_d)$.

**Proof.** Let $M_1$ be the set of tuples that should appear in $T_1 \bowtie T_2$ by the definition of natural join:

$T_1 \bowtie T_2$ contains all and only tuples $t$ such that $t[T_1] \in T_1$ and $t[T_2] \in T_2$, where $t[T_1]$ ($t[T_2]$) is a tuple obtained from $t$ by keeping only its attributes in $T_1$ ($T_2$).

Let $M_2$ be the set of tuples retrieved by $\Pi_S(\sigma_P(T_1 \times T_2))$. We will prove $M_1 \subseteq M_2$ and $M_2 \subseteq M_1$, which will establish the fact that $M_1 = M_2$.

**Proof of $M_1 \subseteq M_2$.** Consider any tuple $t \in T_1 \bowtie T_2$. We will show that $t \in \Pi_S(\sigma_P(T_1 \times T_2))$. For this purpose, let $t_1 = t[T_1]$ and $t_2 = t[T_2]$. By the definition of natural join, we know that $t_1 \in T_1$ and $t_2 \in T_2$. Hence:

$$(t_1, t_2) \in T_1 \times T_2$$

where $(t_1, t_2)$ is the tuple obtained by concatenating $t_1$ and $t_2$. Since $t_1.A_i = t_2.A_i$ for each $i \in [1, d]$, we know

$$(t_1, t_2) \in \sigma_P(T_1 \times T_2).$$

Finally, as $t$ is obtained from $(t_1, t_2)$ by discarding $t_2.A_1, ..., t_2.A_d$, we have:

$$t \in \Pi_S(\sigma_P(T_1 \times T_2)).$$

**Proof of $M_2 \subseteq M_1$.** Every tuple in $\Pi_S(\sigma_P(T_1 \times T_2))$ is produced from the following process. First, fix a tuple $t_1 \in T_1$ and a tuple $t_2 \in T_2$. Concatenate them to obtain a tuple $(t_1, t_2) \in T_1 \times T_2$. Check whether $t_1.A_i = t_2.A_i$ for every $i \in [1, d]$. If yes, we generate a tuple $t$ from $(t_1, t_2)$ by discarding $t_2.A_1, ..., t_2.A_d$.

It suffices to prove that $t[T_1] \in T_1$ and $t[T_2] \in T_2$, namely, $t \in T_1 \bowtie T_2$ by definition of natural join. This is obvious because $t_1 = t[T_1]$ and $t_2 = t[T_2]$. 

1